Microbiota and diet in infants and young children

Claudia Leong

A thesis submitted for the degree of

Doctor of Philosophy

at the University of Otago, Dunedin,

New Zealand

2019

Abstract

Diet is known to modify the composition of the gut microbiota. However, few studies have been conducted to assess the relationship between diet and the gut microbiota in infants during the complementary feeding period, or in young children. Associated with the rapid advancement of technology in the gut microbiota field, several bioinformatic data handling processes have been developed to be used to identify links between diet and the gut microbiota. While many common statistical methods have been used, the most appropriate method to determine associations between dietary components and microbiota composition is open to debate.

The overall aim of this thesis was to examine the effects of diet on the composition of the gut microbiota in children using appropriate nutritional, microbiological, and statistical methods, by answering three key questions:

1. How is diet during the complementary feeding period associated with children's subsequent gut microbiota composition at 12 months?
2. Can a food frequency questionnaire (FFQ) measure intake of dietary components thought to influence the composition of the gut microbiota?
3. In what ways is diet associated with the composition of the gut microbiota at 5 years of age?

Chapters 3 and 4 examine the effect of feeding method at 7 months of age on the composition of the gut microbiota at 12 months. The composition of the gut microbiota in 74 participants participating in a randomised controlled trial (the 'Baby-Led Introduction to SolidS (BLISS) study') was determined by sequencing 16S rRNA genes, and 3-day weighed diet records (WDR) were used to estimate dietary intake. In a novel approach to analysing gut microbiota data, mediation models were used to demonstrate that 29% and 25% of the link between feeding method (BLISS vs Control) and alpha diversity at 12 months could be explained by lower 'fruit, vegetables, nuts and legumes' and dietary fibre intake in the BLISS group at 7 months.

Chapter 5 focuses on determining the relative validity and reproducibility of an FFQ for assessing amount and ranking of intakes of nutrients and foods that have been reported to influence the composition of the gut microbiota. One hundred parent-child pairs completed a 3-day WDR and an FFQ on two separate occasions four weeks apart.

The FFQ was found to have acceptable validity for ranking intakes of energy, carbohydrate, dietary fibre, total non-starch polysaccharides (NSP), and insoluble NSP, when compared with the 3-day WDR, and very good reproducibility when measured over four weeks. The FFQ was also suitable for assessing mean absolute intake of carbohydrate, dietary fibre, and total NSP.

Chapters 6 and 7 examine the relationship between diet (measured using the validated FFQ) and gut microbiota composition (determined by sequencing 16S rRNA genes) at 5 years of age. Data from 319 participants from the 'Prevention of Overweight in Infancy (POI) study' were available at 5 years of age. Using compositional principal component analysis (PCA), 3 gut microbiota profiles were identified. Profile 2 (positive loadings on *Bacteroides*; negative loadings on uncultured Christensenellaceae and Ruminococcaceae) and profile 3 (positive loadings on *Faecalibacterium*, *Eubacterium* and *Roseburia*) were independently associated with body mass index (BMI) z-score, and dietary components (fibre, total NSP, and 'meat, fish, poultry' intake), respectively.

These results show that in older infants and young children, certain dietary components ('fruit and vegetables', fibre, total NSP and 'meat, fish, poultry') and BMI z-scores were associated with aspects of the gut microbiota (alpha diversity and gut microbiota profile scores). Collectively, the studies in this thesis demonstrate the utility of a validated FFQ as a dietary assessment tool for use in large studies assessing diet and the gut microbiota, and the usefulness of two different statistical methods (mediation, and compositional PCA) to determine linkages between food components and the gut microbiota. These robust methods further our understanding of the relationships between diet and the gut microbiota in infants and young children.

Preface

The Candidate's supervisory team consisted of the following:

University of Otago

Department of Human Nutrition:

Assoc. Prof. Anne-Louise Heath (ALH)

Dr. Jillian Haszard (JH)

Department of Medicine:

Prof. Rachael Taylor (RT)

Department of Microbiology and Immunology:

Prof. Gerald Tannock (GT)

Deakin University (Melbourne)

School of Exercise & Nutrition Sciences:

Dr. Ewa Szymlek-Gay (ESG)

The Candidate's PhD convenor was:

University of Otago

Department of Human Nutrition:

Assoc. Prof. Rachel Brown (RB)

In addition, the Candidate also received technical research advice from:

University of Otago

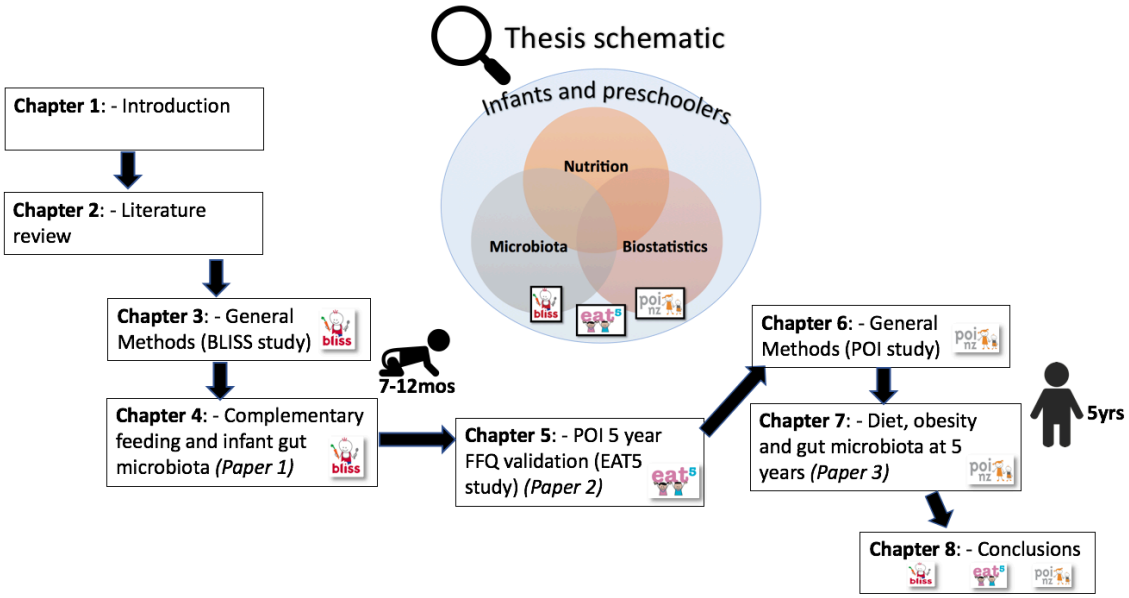
Department of Human Nutrition:

Elizabeth Fleming (EF)

Department of Microbiology and Immunology:

Blair Lawley (BL)

This thesis includes three research projects as summarized in the thesis schematic shown here:



The Baby-Led Introduction to SolidS (BLISS) study is a randomised controlled trial (RCT) funded by Lotteries Health Research, Meat & Livestock Australia, Karitane Products Society, Perpetual Trustees, New Zealand Federation of Women's Institutes, and the University of Otago. ALH and RT were the co-Principal Investigators of the BLISS study and were responsible for securing funds and designing the studies. Recruitment was from November 2012 to February 2014 and data collection by the BLISS research team was from November 2012 to April 2016.

The Eating Assessment in Toddlers 5 (EAT5) study involved the collection of 3-day weighed diet records (WDR) from 5-year-old pre-schoolers. Collection of these data was primarily by five Master of Dietetics students. Each of the 5 masters students collected approximately 20 WDRs and food frequency questionnaires (FFQ) before and after collecting the weighed diet records. The EAT5 study data collection started in March 2015 and ended in December 2017. The EAT5 FFQ consisting of 123 food items was developed by ALH and RT, based on previous EAT study FFQs, and with input from Dr Sonya Cameron (University of Otago).

The Prevention of Overweight in Infancy (POI) intervention study is a large RCT funded by the Health Research Council of New Zealand. RT was the co-Principal Investigator of the POI study (with Professor Barry Taylor, University of Otago), and ALH and GT were co-Investigators for the POI study. The POI investigator team was

responsible for securing funds and designing the study. The POI study started in May 2009 and data collection ended in April 2017.

Kai-culator was the dietary assessment software programme used to input and calculate dietary data from the WDRs in the BLISS and EAT5 studies. Technical advice for Kai-culator was obtained from EF. All statistical analyses were conducted by the Candidate with advice from JH. Microbiota analyses were carried out by BL. Stata 13 (StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP) was the statistical software programme used to conduct most of the analyses in this thesis. STAMP (Parks et al., 2014. *Software Package: Version 2.1.3*. GNU General Public License) was one of the bioinformatics software programmes used to generate some microbiota graphs. 'R' (R Core Team. 2013. *R Foundation for Statistical Computing*, Vienna, Austria) is a statistical and bioinformatics software programme that was also used in the later chapters of this thesis. GraphPad Prism (version 8; *GraphPad Software*, La Jolla California, USA) was a scientific software programme used to create some of the graphs used in this thesis.

The Candidate was responsible for the following:

Data collection:

- Carried out EAT5 study measurements for 1 participant, including anthropometric data, explanation of the completion of the weighed diet record and conducting the food frequency questionnaire interview twice.

Dietary data:

- Developed food groups for BLISS (n = 9), EAT5 (n = 12) and POI (n = 12) studies.
- Allocated the BLISS weighed diet record items (n = 1682) into the new food groups for input into Kai-culator (Appendix C).
- Developed protocols and codebooks for the data entry and checking of WDRs for the BLISS and EAT5 studies (Appendix H).
- Developed protocols and codebooks for the data entry and checking of FFQs for the EAT5 and POI studies (Appendix I and Appendix J).
- Carried out data entry of all 199 EAT5 and 546 POI5 FFQs into an online platform 'ffq.otago.ac.nz'.

- Conducted thorough checks on each nutrient line, and developed and assisted in final decisions on nutrient lines for the EAT5 and POI5 FFQs (Appendix K).
- Developed and checked nutrient lines for fibre fractions such as non-starch polysaccharides, resistant starch, pectin, cellulose and hemicellulose (Appendix L and Appendix M).
- Checked output of all EAT5 and POI5 FFQs.
- Entered 1 EAT5 WDR and checked and corrected all 100 EAT5 WDRs.
- Coded individual food items (n = 1010) from the 100 WDRs to the 12 food groups developed for the EAT5 study.
- Conducted thorough checks on the exported dietary data from Kai-culator for inconsistencies.

Gut microbiota data:

- Carried out data entry for all the microbiota questionnaires for the BLISS study.
- Undertook data exploration using STAMP and 'R' (Appendix N).

Statistical analysis:

As the PhD is based on secondary data analyses of data collected from 3 studies, there was a strong focus on designing statistical analysis plans before undertaking any statistical analyses.

- Candidate worked closely with JH, ALH and RT to prepare the first and subsequent drafts of the statistical analysis plans for Chapters 4, 5 and 7.
- Prepared a summary of lessons learnt and limitations of the statistical analysis plan for Chapters 4, 5 and 7 (Appendix E, Appendix F and Appendix N).
- Conducted data cleaning, combining and coding of all variables (dietary, gut microbiota and demography) for the BLISS, EAT5 and POI studies.
- Carried out the statistical analyses in Chapters 4, 5 and 7 with guidance from JH.

Figures and tables:

- Prepared the tables and figures in Chapters 4, 5 and 7 with guidance from supervisory team.
- Designed the figures in Chapter 2 from the Candidate's own understanding of the literature.

- Generated additional figures in Chapters 4 and 7 from Candidate's exploration of the software programs 'R', STAMP, and GraphPad Prism.

Publications:

- Prepared the first and subsequent drafts of the following published papers from Chapters 4 and 5.

Leong, C., Haszard, J. J., Lawley, B., Otal, A., Taylor, R. W., Szymlek-Gay, E. A., Fleming, E. A., Daniels, L., Fangupo, L. J., Tannock, G. W., Heath, A.-L. M. (2018). Mediation analysis as a means of identifying dietary components that affect the fecal microbiota of infants weaned by modified baby-led, compared to traditional, approaches. *Applied and Environmental Microbiology, 84*(18), e00914-00918. doi:10.1128/aem.00914-18

Leong, C., Taylor, R. W., Haszard, J. J., Fleming, E., Tannock, G. W., Szymlek-Gay, E. A., Cameron, S. L., Yu, R., Carter, H., Chee, L. K., Kennedy, L., Moore, R., Heath, A.-L. M. (2018). Relative validity and reproducibility of a food frequency questionnaire to assess nutrients and food groups of relevance to the gut microbiota in young children. *Nutrients, 10*(11), 1627. doi: 10.3390/nu10111627

- Prepared the first and subsequent drafts of manuscripts arising from Chapter 7.

Leong, C., Haszard, J. J., Heath, A.-L. M., Tannock, G. W., Lawley, B., Cameron S. L., Szymlek-Gay, E. A., Gray, A. R., Taylor, B. J., Galland, B. C., Lawrence, J. A., Otal, A., Hughes, A., Taylor, R. W. (2019). Using compositional principal component analysis to describe children's gut microbiota profiles in relation to diet and body composition. *(Submitted The American Journal of Clinical Nutrition)*

Co-authored publications:

- Dietary data from the BLISS study in Chapter 4 of this thesis were used. The Candidate was responsible for the allocation of individual food items into the food groups presented in this publication.

Williams-Erickson, L., Taylor, R. W., Haszard, J. J., Fleming, E., Daniels, L., Morison, B., **Leong, C.**, Fangupo, L. J., Wheeler, B. J., Taylor, B. J., Te Morenga, L., McLean, R. M., Heath, A.-L. M. (2018). Impact of a modified version of baby-led weaning on infant food and nutrient intakes: the BLISS randomized controlled trial. *Nutrients, 10*(6), 740. doi:10.3390/nu10060740

- Dietary data from the EAT5 study in Chapter 5 of this thesis were used. The Candidate was responsible for the data obtained from WDRs and FFQs presented in this publication.

Fangupo, L. J., Haszard, J. J., **Leong, C.**, Heath, A.-L. M., Fleming, E., Taylor, R. W. (2019). Relative validity and reproducibility of a food frequency questionnaire to assess energy intake from minimally processed and ultra-processed foods in young children. *Nutrients, 11(6),* 1290. doi: 10.3390/nu11061290

Conference oral presentations and posters:

- Presented information from Chapter 2 at a conference in New Zealand (November 2018).

Leong, C., Heath, A.-L. M., Haszard, J. J., Tannock, G. W., Lawley, B., Szymlek-Gay, E. A., Taylor, R. W. (2018). Understanding diet, gut microbiota and their interrelationship with obesity from complementary feeding to start of school: evidence to guide future research. *Postgraduate and Early Career Nutrition conference* (Auckland, NZ). [Abstract for oral presentation]

- Presented results from Chapter 4 at a conference and symposium in New Zealand (August and September 2017).

Leong, C., Taylor, R. W., Tannock, G. W., Haszard, J. J., Lawley, B., Szymlek-Gay, E. A., Heath, A.-L. M. (2018). Effects of a 'Baby-led' Approach to Complementary Feeding on Infant Gut Microbiota – A randomised controlled trial. *27th Annual Queenstown Molecular Biology (QMB) Meeting* (Queenstown, NZ). [Abstract for oral presentation]

Leong, C., Taylor, R. W., Tannock, G. W., Haszard, J. J., Lawley, B., Szymlek-Gay, E. A., Heath, A.-L. M. (2018). Effects of a 'Baby-led' Approach to Complementary Feeding on Infant Gut Microbiota: the Baby-Led Introduction to SolidS (BLISS) randomised controlled trial. *University of Otago Student Research Symposium Te Wānaka Rakahau - Ākoka 2017* (Dunedin, NZ). [Abstract for oral presentation]

- Presented results from Chapter 5 at a symposium in New Zealand (February 2019).

Leong, C., Taylor, R. W., Haszard, J. J., Fleming, E., Tannock, G. W., Szymlek-Gay, E. A., Cameron, S. L., Yu, R., Carter, H., Chee, L. K., Kennedy, L., Moore, R., Heath, A.-L. M. (2019). Validation of a Food Frequency Questionnaire to Assess Intake of Fibre and Fibre-Related Food Groups of Relevance to the Gut Microbiota. *Focus on Fibre and Food Monitoring symposium* (Dunedin, NZ). [Abstract for oral and poster presentation]

- Submitted abstract of results from Chapter 7 for international conferences in China (September 2019) and Ireland (October 2019).

Leong, C., Haszard, J. J., Heath, A.-L. M., Tannock, G. W., Lawley, B., Cameron S. L., Szymlek-Gay, E. A., Gray, A. R., Taylor, B. J., Galland, B. C., Lawrence, J. A., Otal, A., Hughes, A., Taylor, R. W. (2019). Describing gut microbiota profiles and their associations with children's diet and body composition. *11th Asia Pacific Conference on Clinical Nutrition* (Nanjing, China). [Abstract for oral presentation]

Leong, C., Haszard, J. J., Heath, A.-L. M., Tannock, G. W., Lawley, B., Cameron S. L., Szymlek-Gay, E. A., Gray, A. R., Taylor, B. J., Galland, B. C., Lawrence, J. A., Otal, A., Hughes, A., Taylor, R. W. (2019). Compositional principal component analysis generates gut microbiota profiles that associate with children's diet and body composition. *13th European Nutrition Conference, Federation of European Nutrition Societies (FENS)* (Dublin, Ireland). [Abstract for poster presentation]

Acknowledgements

This PhD journey has been amazing. As I reflect upon my present blessings, the words here cannot fully express the thankfulness I have for everyone, even those not listed here, who have played a part in my life journey.

I am really blessed to have 5 awesome supervisors. Associate Professor Anne-Louise Heath, for forming my wonderful team of supervisors. I appreciate your guidance, structured meetings, and detailed feedback. You always make time to listen to my problems, both in my studies and personal life. Thank you for being so kind and loving. Professor Rachael Taylor, for being so supportive and always making time for our weekly meetings. I appreciate your quick feedback, encouraging words, and expert advice. Dr. Jillian Haszard, for your valuable ideas. Thank you for being so patient in explaining the various concepts and for being such a great teacher. Professor Gerald Tannock, for your constructive feedback. I appreciate the many questions you bring up during our meetings that help guide my studies. Dr. Ewa Szymlek-Gay, for your advice and support, and setting aside time for the zoom meetings from overseas. I am truly honoured to be supervised by such inspiring role models.

Thank you to my 2 technical advisors. Dr. Blair Lawley, for teaching me about the methods in microbiology. Thank you for being so patient with my incessant questions. Elizabeth Fleming, for teaching me about the usage of Kai-culator. You have been so kind with your time, helping me in my projects, and it was a great learning experience working with you to solve the technical challenges.

My convenor, Associate Professor Rachel Brown, thank you for always having your door open to hearing my stories, and I really appreciate your valuable life advice. Thank you for your cheerfulness in handling all the paper work from my 3 different departments and an overseas supervisor.

Sincerely, and from the bottom of my heart, I would like to thank everyone from the BLISS, EAT5 and POI research teams. Thank you for all the hard work in completing the data collection. Thank you to the participants and their families from the BLISS, EAT5 and POI studies. Without both the research teams and participants, my PhD would not have been possible.

Thank you to staff and postgraduate students from the Human Nutrition, Medicine, and Microbiology and Immunology departments, both past and present. Thank you, Madeline, Anne, Frances, Shae and Clare for the administrative help, Gorden for the IT support, Charlie for the Kai-culator support, Gavin and Kim for the online FFQ support, and Alan for the microbiology laboratory insight. Thank you, Associate Professor Lisa Houghton for your support and great ideas. Thank you, Kieran, Karl, Michelle, Ash, Ken and Ivy, for your friendship, chats, table tennis games, and lovely cakes. My fellow postgraduates, for your friendship and support. Lisa Daniels, Pouya Saeedi, and Ruth Harvie, for your kindness, support and sharing your valuable knowledge. Chaya, Alex, Sarahmarie, Maggie, Caleb, Evie, Karin, Alice, Melanie, Silke and Lara, I enjoyed our tea time chats, intellectual discussions, lunch and dinner outings, the ski trip and the escape room game! I am glad to be able to go through this journey together with you.

Aly Diana, I am blessed to have met such an amazing friend. Thank you for always being around for a chat, for all the crazy adventures together during the time in Dunedin, and for being like family. Ros, Ian, Lika and Jason, who I got to know through Aly, thank you for your warm hospitality, delicious cooking and exciting hiking trips. Those times together have become memories etched to my life. Annisa and Lucia, I am thankful that we were able to start our PhD journeys at the same time. Thank you for your friendship, encouragement, laughter, fun times and lovely lunch dates.

My ex-bosses from Singapore, Professor Jeya Henry and Dr. Ciarán Forde, for your care and guidance. Thank you for your kind encouragement and messages. My ex-colleagues from Singapore, Rina, Viren, Janice, Feng Ming, Bhupi, Aiting, Wani, and Edwin, thank you for your friendship and keeping in touch even when we are far apart. Thank you for the phone calls, messages and meet ups whenever possible. Dr. Agnes Tey, I am fortunate to have such a loving and generous friend. Thank you, for the introduction to the Department of Human Nutrition.

My undergraduate supervisors, seniors and friends from Massey University, for being such great friends. Dr. Lara Matia-Merino and Associate Professor Kelvin Goh, for being my first link to scientific research, and being such great supervisors and

mentors. Wensheng, for your time and wise advice when I first came to Dunedin. Liangjue, Jiehong, and Eileen, for your kind hospitality in the North Island.

A big thank you to my friends in Singapore and the new ones I made in New Zealand. To the members and committees who I served with at Speakeasy Toastmasters Club, Sunshine Women Institutes, and Graduate Women Otago from 2016 to 2019. I appreciate the time learning from you and serving the Dunedin community together. I have learnt so much more of your culture and thank you for making me feel so comfortable here in Dunedin.

A special thank you to my dad, my older sister, Dorcas, and brothers, Gaius and Boaz. Thank you for being such a delight, my niece, Jael, and nephews Elijah, Noah and Jedidiah. In remembrance of my dear mother, who was my main motivation to start studying nutrition. My lovely family, you always have a special place in my heart.

Brent, you have been such a sweet surprise. Thank you for being such a strong supporter and always offering a listening ear. You have made such a difference in my life in the past few years.

Last but not least, the University of Otago, thank you for supporting my studies with a PhD scholarship. Thank you for the many courses and training workshops that you made available. I am forever grateful for this life-changing opportunity.

Table of contents

List of tables

List of figures

Abbreviations

16S rRNA	16S ribosomal ribonucleic acid
AOAC	Association of Official Analytical Chemists
ASA	American Statistical Association
BH	Benjamini-Hochberg
BLISS	Baby-Led Introduction to SolidS
BLW	Baby-Led Weaning
BMI	Body mass index
CI	Confidence interval
DNA	Deoxyribonucleic acid
DXA	Dual-energy x-ray absorptiometry
EAT5	Eating Assessment in Toddlers at 5 years
F/B ratio	'Firmicutes to Bacteroidetes' ratio
FCDB	Food composition databases
FDR	False discovery rate
FFQ	Food frequency questionnaire
FMI	Fat mass index
FV-adjusted	Fruit and vegetables adjusted
HMO	Human milk oligosaccharide
HMP	Human Microbiome Project
HTS	High throughput sequencing
IBCLC	International Board-Certified Lactation Consultant
MoH	Ministry of Health
MSM	Multiple source method
NCBI	National Center for Biotechnology Information
NSP	Non-starch polysaccharides
NZDep	New Zealand Index of Deprivation
OTU	Operational taxonomic unit
PCA	Principal component analysis
PCoA	Principal co-ordinate analysis
PCR	Polymerase chain reaction
POI	Prevention of Overweight in Infancy
RCT	Randomised controlled trial
SCFA	Short chain fatty acid
SD	Standard deviation
VS	versus
WDR	Weighed diet record
WHO	World Health Organisation
V4	Variable region 4

1 Introduction

Habitual diet influences the establishment and composition of the gut microbiota. The food we eat is not only important to us, the human hosts, but also feeds the trillions of microbes living in our gut, called the gut microbiota. The interplay between host, gut microbiota and diet is intricate and needs to be untangled so that impacts on human health can be elucidated.

It is recognized that a critical window of rapid change in the gut microbiota exists between the end of solely milk feeding (i.e. the introduction of solids) at around 4 - 6 months of age (Fallani et al., 2011), until around 3 years of age, when the gut microbiota is considered to have developed into a more adult-like composition (Laursen, Bahl, Michaelsen, & Licht, 2017). Existing studies in children demonstrate that, for example, high fibre diets are associated with significant enrichment of Bacteroidetes (De Filippo et al., 2010), and that dairy intake is positively associated with the 'Firmicutes to Bacteroidetes' (F/B) ratio (Smith-Brown, Morrison, Krause, & Davies, 2016). However, few studies have looked at the effect of introducing solid foods (also known as complementary feeding) on the gut microbiota (Laursen et al., 2017), and none have looked at the impact of 'Baby-Led Weaning (BLW)'.

Traditionally, parents have been encouraged to start spoon-feeding their infant puréed foods from around 6 months of age, progressing to mashed, then chopped foods with the aim that they will be eating family foods by around 12 months of age (Ministry of Health, 2008). However, an alternative method of complementary feeding, known as BLW, is becoming popular in New Zealand (Morison et al., 2016), the United Kingdom (Brown & Lee, 2011), the United States (Beal, 2016) and Canada (D'Andrea, Jenkins, Mathews, & Roebothan, 2016). In BLW, the infant feeds themselves whole pieces of family food from the start of complementary feeding, often eating with the family, so that they would be expected to be eating a more adult type-diet at a much earlier age than traditionally spoon-fed infants. The impact of a modified version of BLW on growth (Taylor et al., 2017b), choking risk (Fangupo et al., 2016), and iron status (Daniels et al., 2018) have been investigated. However, no studies so far have studied the effect of a baby-led approach to infant feeding on the composition of the gut microbiota.

Obesity is on the rise in young children, adolescents and adults in both developing and developed populations (The GBD Obesity Collaboration et al., 2014). A rapidly expanding literature suggests that our gut microbiota may have both beneficial and harmful impacts on our health, with human studies reporting associations between particular characteristics of the gut microbiota and a wide range of health conditions including: obesity, diabetes, inflammatory bowel disease, irritable bowel syndrome, and allergies (Bäckhed, Ley, Sonnenburg, Peterson, & Gordon, 2005; Baothman, Zamzami, Taher, Abubaker, & Abu-Farha, 2016; Honda & Littman, 2012; Ringel & Carroll, 2009). Dysbiosis of the gut microbiota community has been linked to obesity in adults (Bäckhed et al., 2012; Valdes, Walter, Segal, & Spector, 2018). However, less is known about any relationship with childhood obesity (Taylor, 2016). Moreover, because childhood gut microbiota has different composition, less stability and less diversity than an adult's gut microbiota (Voreades, Kozil, & Weir, 2014), associations between the gut microbiota and health outcomes such as obesity, may be different in growing children to those in adults.

To date, just 14 studies have been carried out examining associations between diet and the composition of the gut microbiota in older infants and younger children (from the start of complementary feeding to 11 years of age). Dietary data in these studies were obtained using 24-hour recalls (Smith-Brown et al., 2016), diet records (Amarri et al., 2006; De Filippo et al., 2017; Krebs et al., 2013; Qasem et al., 2017) or food frequency questionnaires (FFQ) (La-Ongkham, Nakphaichit, Leelavatcharamas, Keawsompong, & Nitisinprasert, 2015; Laursen et al., 2016; Nakayama et al., 2015; Nakayama et al., 2017; Smith-Brown et al., 2016; Zhong et al., 2019). Diet records and 24-hour recalls are widely accepted dietary assessment methods but have high participant and researcher burden and may not represent the variety of foods consumed by the child over a longer period of time. Food frequency questionnaires have the advantage that they have much lower respondent and researcher burden so can be used in large studies. However, the FFQ must have been validated for the nutrient(s) or food group(s) of interest in the population under study (Willett, 1998). However, no studies have validated an FFQ specifically designed to look at nutrients and food groups that would be expected to influence the composition of the gut microbiota.

This emerging research area of the impact of childhood diet on the gut microbiota has its challenges, not only in getting good quality dietary data for infants and children,

but also in interpreting gut microbiota data. Gut microbiota data can be challenging especially because they are complex, compositional (Quinn, Erb, Richardson, & Crowley, 2018), highly dimensional (Li, 2015), and include many zeros (Xia, Sun, & Chen, 2018). Many bioinformatic data handling processes (Dhariwal et al., 2017; Kuczynski et al., 2011; Parks, Tyson, Hugenholtz, & Beiko, 2014; Vázquez-Baeza, Pirrung, Gonzalez, & Knight, 2013) have been developed in recent years to attempt to make the statistical analysis of data involving the gut microbiota accessible to researchers. However, this in a certain sense, has resulted in a disconnect between the underlying science and the statistical analysis of data, as researchers may be limited to using pre-specified statistical analysis packages with particular outputs. As such, there is still no consensus on the most appropriate methods of statistical analysis (Knight et al., 2018; Pollock, Glendinning, Wisedchanwet, & Watson, 2018; Zhang, Nieuwdorp, Groen, & Zwinderman, 2019).

Therefore, more studies are needed that investigate how the diet affects the composition of the gut microbiota in older infants and young children, with special attention paid to the most appropriate ways to measure dietary intake, and analyse both diet and gut microbiota data.

The overall aim of this thesis was to examine the effects of diet on the composition of the gut microbiota in older infants and young children using appropriate nutritional, microbiological, and statistical methods. The specific objectives were to:

1. Determine associations between diet during the complementary feeding period and the infants' subsequent gut microbiota composition at 12 months (**Chapter 4**).
2. Determine the relative validity of an FFQ for measuring intake of dietary components that have been thought to influence the composition of the gut microbiota in children (**Chapter 5**).
3. Determine associations between diet and the composition of the gut microbiota at 5 years of age (**Chapter 7**).

A summary of the chapters included in this thesis is outlined below:

Chapter 2 introduces complementary feeding (including BLW) and the gut microbiota, evaluates the methods commonly used to measure the diet and the gut microbiota, and discusses the factors that have been reported to be associated with the

composition of the gut microbiota. Lastly, studies looking at the associations of diet on the composition of the gut microbiota in children are discussed.

Chapter 3 gives an overview of the Baby-Led Introduction to SolidS (BLISS) randomised controlled trial (RCT). The BLISS intervention is described together with the methods of data collection.

Chapter 4 reports the associations between diet and the composition of the gut microbiota found in infants during the complementary feeding period in the BLISS study. The application and interpretation of specific statistical analyses (in particular, mediation models) are described in detail (**objective 1**).

Chapter 5 describes the Eating Assessment in Toddlers at 5 years (EAT5) FFQ validation study and presents the results for the validity and reproducibility of the EAT5 FFQ when compared to a 3-day WDR. The EAT5 FFQ is the dietary assessment method used to determine food and nutrient intake in **Chapter 7** (**objective 2**).

Chapter 6 gives an overview of the Prevention of Overweight in Infancy (POI) RCT. The POI interventions are described together with the methods of data collection.

Chapter 7 reports the associations between diet and the composition of the gut microbiota found in young children in the POI study. The application and interpretation of specific statistical analyses (in particular, compositional principal component analysis (PCA) and regression models) are described in detail (**objective 3**).

Chapter 8 concludes the thesis, discusses the implications of the findings, and provides recommendations for future research.

2 Literature Review

Since the introduction of the Human Microbiome Project (HMP) in 2007 (Turnbaugh et al., 2007), the number of studies looking at the gut microbiota has grown rapidly. Moreover, numerous studies have recognized diet as a potentially modifiable factor that can influence the gut microbiota (David et al., 2014; Wu et al., 2011; Zmora, Suez, & Elinav, 2019). However, many of these papers are based on studies in animals and adult humans, and few studies have looked at the effect of diet on the gut microbiota in infants and young children. In addition, although many statistical methods have been proposed to analyse the relationships between diet and gut microbiota, there are challenges with the interpretation of results and there has been no strong consensus yet on which methods are the most appropriate.

This review evaluates the literature in three main topic areas:

1. How is the diet measured and what are some of the limitations of dietary assessment in infants and children? (sections 2.2, 2.3 and 2.4)
2. How is the gut microbiota measured and what are some of the limitations of gut microbiota analyses? (sections 2.5, 2.6, 2.7 and 2.8)
3. How are the diet and gut microbiota related and what are some of the limitations of the current methods for analysing associations? (sections 2.9, 2.10 and 2.11)

2.1 Search strategy

Studies were identified using the electronic databases MEDLINE (1996-30 April 2019) and SCOPUS (1970-30 April 2019). Articles were only included if they were published in English with the main focus on human studies, although some studies in animals were considered. **Table 2.1** outlines the search strategies and key terms used. Relevant references were also identified from the reference lists of the articles retrieved.

Table 2.1: Search strategies and terms used within this literature review.

Search terms used in the dietary section

1) Breastmilk or breast milk
2) Breastfeeding or breast feeding
3) infant milk
4) infant formula
5) complementary feeding
6) baby-led weaning or baby led weaning
7) dietary assessment
8) food frequency questionnaire
9) (8) AND validation
10) (8) AND (reproducibility OR repeatability)

Search terms used in the gut microbiota section

1) gut micro*
2) (1) AND host health
3) (1) AND short chain fatty acid
4) (1) AND dysbiosis
5) (1) AND composition
6) (1) AND development
7) (1) AND compositional data
8) (1) AND features
9) (1) AND clustering
10) (1) AND pre-pregnancy BMI
11) (1) AND mode of delivery
12) (1) AND (parity OR siblings)
13) (1) AND (feeding type OR infant feeding)
14) (1) AND antibiotics
15) 16S rRNA
16) high throughput sequencing
17) alpha diversity
18) beta diversity
19) bacterial taxa

Search terms used in the diet and gut microbiota association section

1) gut micro*
2) diet
3) food
4) nutrition
5) nutrient
6) environment
7) (2) OR (3) OR (4) OR (5) OR (6)
8) (1) AND (7)
9) obese OR obesity
10) overweight
11) BMI
12) (9) OR (10) OR (11)
13) (1) AND (7) AND (12)
14) limit (8) to (English language and humans and ‘all child (0 to 18 years)’)
15) limit (13) to (English language and humans and ‘all child (0 to 18 years)’)

2.2 Introduction to infant milk and complementary feeding in children

From birth until around 4 to 6 months, milk (breast milk or infant formula) is the main nutrient source for healthy term infants (Ministry of Health, 2008). After that, other foods are introduced in addition to milk; this is referred to as complementary feeding.

2.2.1 Breast milk and infant formula

Exclusive breastfeeding until the infant is around 6 months of age is recommended by both the New Zealand Ministry of Health (NZ MoH) (Ministry of Health, 2008) and the World Health Organization (WHO) (World Health Organization, 2019). Exclusive breastfeeding is defined by the WHO as only breast milk, from either the breast or expressed, and prescribed medications given from birth (World Health Organization, 2019). If the infant is not given breast milk, or is partially breastfed, they should be given infant formula to at least 1 year of age (Ministry of Health, 2008).

2.2.2 Complementary feeding

Although breastfeeding is recommended to be continued until at least 1 year, or beyond (Ministry of Health, 2008), after 6 months of age, breast milk or infant formula alone are not sufficient to support optimal infant growth and development (Butte, Lopez-Alarcon, & Garza, 2002; WHO, 2002). Hence, solids need to be introduced to complement the milk intake. There are 2 main approaches to complementary feeding: the traditional method of spoon-feeding, and the increasingly popular baby-led weaning method (BLW). **Table 2.2** compares the practices of traditional spoon-feeding and BLW.

Table 2.2: Comparison of traditional spoon-feeding and Baby-led weaning (BLW).

Key points	Traditional spoon-feeding	Baby led weaning
Starting time	4 to 6 months.	6 months.
Spoon-feeding	Infants are spoon-fed at the start of complementary feeding.	Infants are encouraged to feed themselves all of their food from the start of complementary feeding.
Texture of foods	Gradually introduced to solid foods, starting with purées, mashed, chopped (finger foods [a]) and then progressing to family foods.	Introduced to finger foods [a] from the start of complementary feeding. Puréed foods are generally not offered as infants at this age do not have the manual dexterity to use utensils to feed themselves purees.
Family foods	Introduced from 12 to 24 months.	Introduced from start of complementary feeding. Introduced in the form of finger foods [a].
Family meals	Infant may eat separately from the rest of the family as the infant needs to be spoon-fed, making it difficult for the spoon-feeding parent to eat their own food at the same time.	Infant may be more likely to eat together with the family at mealtimes since they are more likely to be eating the same foods.
Infant choice	Infants are not given as much choice about what, how much, and how quickly they eat as they are being spoon-fed by someone else.	Parents/ caregivers choose a range of foods to offer to the infant and the infant chooses which foods to eat, the amount, and the pace that they eat at.

Information from Cameron, Heath, and Taylor (2012a); Ministry of Health (2008); Rapley (2011); Rapley and Murkett (2008); WHO (2002). [a] Finger foods are foods that can be picked up by the infant and eaten with the fingers.

BLW has been growing in popularity compared to the traditional spoon-feeding method for a number of reasons including the proposed benefits of increasing dietary diversity due to the earlier introduction of family foods, decreased 'picky eating', and an increased likelihood of the family sitting together at meals (Brown & Lee, 2013). Concerns about infants following a baby-led approach to infant feeding having a higher risk of iron deficiency and choking have been partially addressed in studies which found that infants following a modified version of BLW did not have a higher risk of iron deficiency (Daniels et al., 2018) or choking (Fangupo et al., 2016) compared to

infants following traditional spoon-feeding, although this version of BLW was specifically modified to address these potential risks.

Increased dietary diversity has been associated with positive nutritional status (as defined by height-for-age z-scores) in children (Arimond & Ruel, 2004). In addition, in a narrative review that looked at studies in both adults and children, it is hypothesized that the more diverse the diet, the more diverse the gut microbiota, and that this is beneficial as it would allow the gut to be more adaptable to perturbations (Heiman & Greenway, 2016). However, there is still no study, to the Candidate's knowledge, that has assessed the impact of BLW on the gut microbiota.

From 12 until 24 months it is recommended that toddlers are mostly fed family foods (with some harder textures being modified) and are introduced to cows' milk or suitable alternatives to drink instead of infant formula (Ministry of Health, 2008). Continuing breastfeeding on demand is encouraged. Moreover, it is expected that from 24 months, children fully transition to a diet that includes family foods only, with breastfeeding continued if the mother and child desire.

2.2.3 Dietary recommendations for children

Optimal transition from a milk-based diet to a diet including family foods is crucial as it coincides with rapid growth and development. In addition, diet plays an important role in child health as a shift in diet towards an increase in intake of energy-dense foods have been found to be one of the fundamental causes of childhood obesity (WHO, 2019). However, little is still known about the relationship between diet and gut microbiota in young children.

The term 'pre-schoolers' is used to refer to children from 2 to 5 years of age, and the term 'children' for children aged 2 to 12 years. There are 4 food groups that the NZ MOH focuses on in their guidelines for pre-school and older children: at least 3 servings of vegetables and 2 servings of fruit; 5 servings of breads and cereals; 2-3 servings of milk and milk products; and 1-2 servings of lean meat, poultry, seafood, eggs, legumes, nuts and seeds (Ministry of Health, 2012). The recommendations are very similar to those for adults, except that in adults, 6 instead of 5 servings of breads and cereals are advised, and to choose mostly whole grains; low- or reduced-fat milk products or alternatives should be chosen; and 2 servings of legumes, nuts and seeds or

at least 1 serving of lean meat, poultry, seafood and eggs should be consumed (Ministry of Health, 2015).

2.3 Measurement of dietary intake

There are three particularly challenging aspects of measuring the dietary intake of individuals in paediatric microbiota studies: assessment of breast milk/infant formula intake, assessment of overall dietary intake, and measurement of the intake of fibre.

2.3.1 Assessment of breast milk and infant formula intake

The infant's first food source is either breast milk or infant formula, or both. There are various methods used to measure breast milk and infant formula.

Infant breast milk intake can be measured using deuterium oxide (a stable isotope method) or by test weighing. The deuterium oxide method is the gold standard method and does not interfere with habitual feeding patterns (International Atomic Energy Agency, 2010). The mother is given an oral dose of deuterium oxide, then the infant's intake of breast milk is measured over a period of 14 days using urine or saliva samples (collected on day 1, 2, 3, 4, 13 and 14) from the mother and infant to measure the disappearance of the isotope from the mother's body and its appearance in the infant's body (International Atomic Energy Agency, 2010). The disadvantages of the method include its expense, the number of contacts required with the participant (although some recent work showed the possibility of reducing the number of days of samples required (Liu et al., 2019b)), and the challenges of collecting sufficient saliva from the infant.

Test weighing is therefore more commonly used in the literature. In this method, the infant is weighed before and directly after a breastfeed to determine their weight gain during the feed, and this is attributed to the breast milk consumed (Dewey, Heinig, Nommsen, & Lonnerdal, 1991b). Some limitations of the test weighing method include insensible water loss from the infant (which may result in an underestimation of intake of 3% to 10% (Butte, Garza, Smith, & Nichols, 1983; Dewey et al., 1991b), depending on ambient temperature); the risk that habitual breastfeeding behaviour may be altered due to the challenges of weighing a hungry infant before feeding, an infant who may be asleep after feeding, and at all hours of the day and night; as well as the challenges of accounting for milk posseting (regurgitation of milk) between feeds.

In addition, infant breast milk intake can be estimated using values for the specific age group from the literature, most commonly total average daily breast milk intake (e.g., 0.78 kg/d), determined using the stable isotope gold standard (da Costa et al., 2010), or an average quantity per breastfeed (e.g., 76 g/breastfeed), determined using the test weighing method (Kent et al., 2006). This estimation of infant breast milk intake is relatively simple, has low respondent burden, and the difference between the measured amount of infant formula and this estimate can be used to approximate breast milk intake for mixed feeders. However, this is just a crude estimation of the amount an individual infant will have actually consumed. Although in the ideal situation the stable isotope method should be used (da Costa et al., 2010), for population studies, the estimated breast milk intake is usually preferred due to its ease of use and its much lower costs. In addition, the test weighing method is often not preferred due to the high respondent burden to the participant because in many studies breast milk intake is not the main outcome, and the participants are already being required to perform many tasks.

Assessing intakes of infant formula is relatively more straightforward, although it is important to take into consideration the observation that many parents do not prepare formula according to the manufacturer's instructions – for example by packing down the formula powder into the measuring spoon, or by diluting the product more than it should be in order to save money on what is an expensive product. Therefore, infant formula intake should be determined by: the parent weighing the amount of formula and the amount of water used to prepare the bottle, the weight of prepared formula offered, and the amount *not* consumed by the infant. It is essential that these separate amounts are weighed rather than using volumes because graduations on formula bottles are often incorrect (Luque et al., 2013).

After measuring or estimating the weight of breast milk or infant formula consumed, the amount of nutrients consumed is usually calculated using food composition databases (e.g., USDA food composition database) (USDA, 2017). Direct sampling and analysis of breast milk is possible, but challenging, because colostrum, transitional milk, and mature milk (foremilk and hind milk) have different composition which needs to be considered (Ballard & Morrow, 2013). Moreover, breast milk composition varies within feeds, diurnally, and with storage (if expressed breast milk is used) (Ballard & Morrow, 2013). Furthermore, different methods are used to measure

the energy and protein content of breast milk (Gidrewicz & Fenton, 2014). Estimating nutrient intake from infant formula is also challenging due to the limited range of infant formulas in food composition databases (e.g., FOODfiles™ (New Zealand Institute for Plant & Food Research Limited & Ministry of Health, 2014)), and the practice of overage, in which manufacturers add higher amounts of unstable nutrients to the product than appears on the label so that the product will have that nutrient value at the end of the shelf life stated on the package (Food Standards Australia New Zealand, 2008).

2.3.2 Assessment of the overall diet

Challenges of assessing dietary intake of young children

Once the infant starts eating solid foods in addition to breast milk or/and infant formula (i.e. complementary feeding), assessing the dietary intake is challenging, and no one method can assess an individual's or group's intake perfectly. The three main methods of dietary assessment used in research are: diet records, 24-hour recalls, and food frequency questionnaires (FFQ). These methods can capture, to a greater or lesser extent, nutrient intake (e.g., dietary fibre), food intake (e.g., bread), food group intake (e.g., cereals), and dietary patterns (e.g., 'adult' type diet). To determine the nutrient content of the foods eaten, food composition databases are required.

There are six key challenges when assessing food intake in infants and children (≤ 5 years of age). First, parents and caregivers have to act as a proxy to record or recall their child's dietary intake, and dietary data become less reliable when there are other caregivers or when infant feeding occurs outside the home environment (Foster, Adamson, Anderson, Barton, & Wrieden, 2009). Second, infants and young children have a great deal of, often messy, leftovers that must be estimated to calculate the portion that was eaten (Gondolf, Tetens, Hills, Michaelsen, & Trolle, 2012). Third, seasonal fruits and vegetables may not be consumed at the time of data collection, which must be taken into account if usual intake rather than recent intake is assessed (Cade, Thompson, Burley, & Warm, 2002). Usual intake captures the food consumed over a longer period of time and hence will be tricky in infants because their diets are changing rapidly. Fourth, low energy reporting may occur if parents are reluctant to report consumption of 'unhealthy' foods, or have trouble weighing foods (Cook, Pryer, & Shetty, 2000). If energy intake is underestimated, then the intakes of other nutrients

and foods are likely to be underestimated too (Livingstone & Black, 2003). Fifth, there is as yet no comprehensive food composition database that includes the range of food components currently thought likely to influence the gut microbiota (see section 2.10). Sixth, although a duplicate diet can be collected if we want to know exactly what a child has eaten (i.e. a replica of the actual food consumed is collected and analysed in the laboratory for nutrient content), this will not be representative of usual intake. Collecting duplicate diets is expensive, and as challenging as it is to collect in adults, it would be even more difficult for school-children to set aside a duplicate of foods they ate outside of the home (Lightowler & Davies, 2002).

Hence, it is important that dietary assessment methods are selected according to the research question, study design and available resources; and that the analysis and interpretation of the data take into account the selected method's limitations. **Table 2.3** provides a summary of some of the strengths and limitations for these three dietary assessment methods when collecting dietary information in children.

Diet records

Diet records can be weighed or estimated, with weighing providing greater accuracy. The participant records the amounts of all foods and beverages (including snacks) eaten by the child, and details of food preparation and brand names for a pre-specified number of days (Gibson, 2005). To get unbiased energy intakes for children, a 7-day record is recommended (Domel, 1997). However, a 3-day record is more practical and less burdensome on both the participant and researcher, and can determine differences in nutrient intake between groups, although not individual intakes (Livingstone & Black, 2003).

24-hour recalls

In 24-hour recalls, participants are asked by a trained researcher (or self-administered by a computer-assisted program) to recall what their child's exact food and drink intake was during the previous 24-hours (or the previous day). Multiple-pass interviewing techniques, with standardized probe questions, are now considered best practice (Gibson, 2005). Portion sizes, and leftovers, have to be recalled which can introduce error.

Table 2.3: Summary of strengths and limitations of the three main methods of dietary assessment used in a research setting.

	Diet records	**24-hour recalls**	**Food frequency questionnaires (FFQ)**
Strengths	Generally considered to be the gold standard for dietary assessment – especially if weighed.	Less respondent burden than diet records, and less data entry burden if computer assistance is used.	Less respondent and data entry burden than diet records and 24-hour recalls.
	Open-ended, hence the researcher may go back to reanalyse or regroup different food items.	Open-ended and a single 24-hour recall can be used for population mean intakes of nutrients.	Very quick to administer, complete, and analyse.
	Able to calculate usual nutrient, food, and food group intake for individuals and groups, and able to generate dietary patterns, if multiple days are collected.	Able to calculate usual group nutrient, food or food group intakes if a second 24-hour recall is collected.	Able to calculate usual group nutrient, food, and food group intakes.
		Multiple 24-hour recalls can be used to determine individual intakes.	Commonly used to determine dietary patterns.
Limitations	Respondents need to be trained to weigh, or estimate, and fill in the records accurately.	24-hour recalls would not usually be used to determine dietary patterns, though they have been used in that way.	Not appropriate to use for individuals, and should only be used for relatively large studies (i.e. hundreds of participants).
	Electronic dietary scales (ideally accurate to 1 g) need to be provided for weighed records, or portion size aids (e.g., photographs) for estimated records.	Relies on memory for accurate recall of foods and portion sizes eaten and left over.	Dependent on the ability of the participant to accurately estimate and report the frequency and amount eaten, and parent may not be present for all eating occasions
	Parent may not be present for all eating occasions and behaviour may change due to burden	Parent may not be present for all eating occasions and also, may not recall socially undesirable foods and overestimate those that are deemed healthy.	Must be validated in population it is used in – which is time-consuming and expensive.

	Diet records	24-hour recalls	Food frequency questionnaires (FFQ)
Comments	The quality of the diet records should be checked when they are received and inaccuracies and missing entries clarified with the respondent.	Protocols and probe questions must be standardised and pre-tested in the population of interest before use.	List of food items chosen is specific to different countries or cuisines and also dependent on the nutrient of interest.
	Should be collected on non-consecutive days to prevent intake from one day influencing intake on the next day and minimise participant fatigue.	If multiple recalls are collected, should be collected on non-consecutive days to prevent intake from one day influencing intake on the next day and minimise participant fatigue.	Essential that the FFQ is validated in the population (country, sex, age) of interest, as the results are entirely dependent on choosing the right foods and beverages for the list.
	Should be collected on both weekend days and week days (which often differ).	If multiple recalls are collected, should be collected on both weekend days and week days (which often differ).	

Food frequency questionnaires

In an FFQ, participants are asked to record the frequency-of-intake of each of a list of food and beverage items, usually over the past year, month(s), or week(s). In a semi-quantitative questionnaire, portion size estimates are added so that not just frequencies but energy and selected nutrients can be derived from the questionnaire. In a quantitative questionnaire, participants are asked to estimate the amount eaten per serve (Cade et al., 2002; Gibson, 2005). Similar to the 24-hour recall, FFQs rely on recall and estimated portion sizes.

Overall, if the resources are available, the first choice for dietary assessment should generally be a weighed diet record. Fisher and colleagues have shown in their study in infants and toddlers that a 24-hour recall overestimates energy, macronutrient and micronutrient intakes compared to a 3-day weighed diet record (WDR) (Fisher et al., 2008). If sufficient resources are not available, and participant burden is a concern (as is often the case in large studies with hundreds of children), a validated FFQ would be a good choice, especially if quantifying the foods consumed is the intended outcome.

Food composition databases

An appropriate food composition database is needed to be able to determine nutrient intake from diet records, 24-hour recalls and FFQs. Examples of food composition databases include FOODfiles in New Zealand (New Zealand Institute for Plant & Food Research Limited & Ministry of Health, 2014), the United States Department of Agriculture (USDA) food composition database in the U.S. (USDA, 2017), the Australian Food Composition Database in Australia (Food Standards Australia New Zealand, 2019), and McCance and Widdowson's Composition of foods integrated dataset (CoFID) in the UK (Public Health England, 2015) (**Figure 2.1**).

Examples of FCDBs	Country	Carbohydrate	Nutrient name in database	AOAC for dietary fibre	Dietary fibre			
					LMWDF	RS	Non-starch polysaccharides	
							Insoluble HMWDF	Soluble HMWDF
USDA food composition	USA	By difference	Fiber, total dietary	985.29 / 991.43		Incl (RS3)	incl	incl
FOODfiles	New Zealand	By sum of available	FIBTG	991.43		Incl (RS3)	incl	incl
			PSACNS	Englyst			incl	incl
			STARES (n=39)	2002.02		Incl		
McCance and Widdowson's CoFID	UK	By sum of available	AOAC fibre	985.29		Incl (RS3)	incl	incl
			NSP	Englyst			incl	Incl
Other AOAC methods commercially available:				2009.01	incl	Incl	incl	Incl
				2011.25	incl	Incl	incl	Incl

Figure 2.1: Examples of Food Composition Databases that report dietary fibre content of foods.

Information sourced from: New Zealand Institute for Plant & Food Research Limited and Ministry of Health (2014); Public Health England (2015); USDA (2017). Abbreviations: FCDBs, Food Composition Databases; Incl, the component is included in the value found in the stated nutrient; LMWDF, low molecular weight fibre; HMWDF, high molecular weight fibre; RS, resistant starch; FIBTG, total fibre; PSACNS, total non-starch polysaccharides; STARES, resistant starch; AOAC, association of official analytical chemists; NSP, non-starch polysaccharides.

Although national food composition databases report nutrient content for thousands of foods, they often do not include the nutrients or food components required in gut microbiota studies. This means that, for example, most composition databases only have fibre separated into insoluble and soluble fibre, and there is little information, if any, on the amount of other microbiota-relevant food components such as pectin, β-glucan, resistant starch, etc. In addition, they do not include data for all foods, so

substitutions commonly have to be made. Although some apps now promise food composition data for huge numbers of brands, these data are derived from manufacturers, or even users, so are not based on representative samples of foods, and have not undergone the rigorous quality control processes used to generate national food composition databases.

As interest in the impact of diet on the gut microbiota grows, we see an increasing number of food composition databases for resistant starch being generated, for example for n = 155 foods (Murphy, Douglass, & Birkett, 2008), n = 54 foods (Landon, Colyer, & Salman, 2012), and n = 25 foods (Liljeberg Elmståhl, 2002). However, these databases remain very small (a food composition database such as FOODfiles has over 2500 foods (New Zealand Institute for Plant & Food Research Limited & Ministry of Health, 2014)), values for many of the food components of interest such as resistant starch will vary according to food production practices in a particular country so may not be appropriate to use in other countries, and some databases require payment for access (such as the Nutrition Coordinating Center Food and Nutrient Database of pectin values (http://www.ncc.umn.edu/food-and-nutrient-database/)). It would be good to see larger open access resources for such data, that include specific information on food production for each product.

2.3.3 Assessment of fibre intake

Dietary fibre intake is of particular importance in relation to the gut microbiota. This is because just as protein, fats and carbohydrates are energy sources for humans, fibre is an energy source for the gut microbiota. Fibre is a subclass of carbohydrates and can be described using the groupings shown in **Figure 2.2**. It is evident that fibre is complex, and due to this complexity (degree of polymerization, sugar types, linkage types etc.) (Cummings & Mann, 2012) and different chemical structures, different types of fibres are utilized by different species of the gut microbiota (Hamaker & Tuncil, 2014; Louis, Flint, & Michel, 2016).

Fibre is quantified in foods using different 'Association of Official Analytical Chemists' (AOAC) methods such as the AOAC 985.29 (total fibre enzymatic-gravimetric method) for total fibre. This is then reported in food composition databases (**Figure 2.1**). A limitation, however, is that there are few databases available that have values for specific fibre components (Westenbrink, Brunt, & van der Kamp, 2013). For

example, to determine the amount of resistant starch (RS) in food, one has to use specific analytical methods such as AOAC 2002.02 and take into consideration how the food is prepared, as for example, freshly boiled hot potatoes (0.59 gRS/ 100g) have a different RS value compared to previously cooked and reheated potato products like hash browns (1.07 gRS/ 100g) (Landon et al., 2012). The AOAC 2002.02 method measures RS2 (native starch granules protected from digestion by the structure of the starch granule) and RS3 (retrograded starch, e.g. cooked then cooled potato, rice or pasta; RS3 forms as it cools) starches. Hence, cooked then cooled potatoes will have greater amounts of RS3 compared to freshly boiled hot potatoes. In addition, there are also updates in the analytical AOAC method based on the updated definition of fibre, and this produces differences in fibre amounts from the current values found in food composition databases (Rainakari, Rita, Putkonen, & Pastell, 2016). For example, FOODfiles™ 2010 uses the Englyst analytical method to quantify fibre, and insoluble non-starch polysaccharides while FOODfiles™ 2014 uses AOAC 991.43. In New Zealand, Plant & Food Research has stopped calculating non-starch polysaccharide values for new food items when they are added to the database.

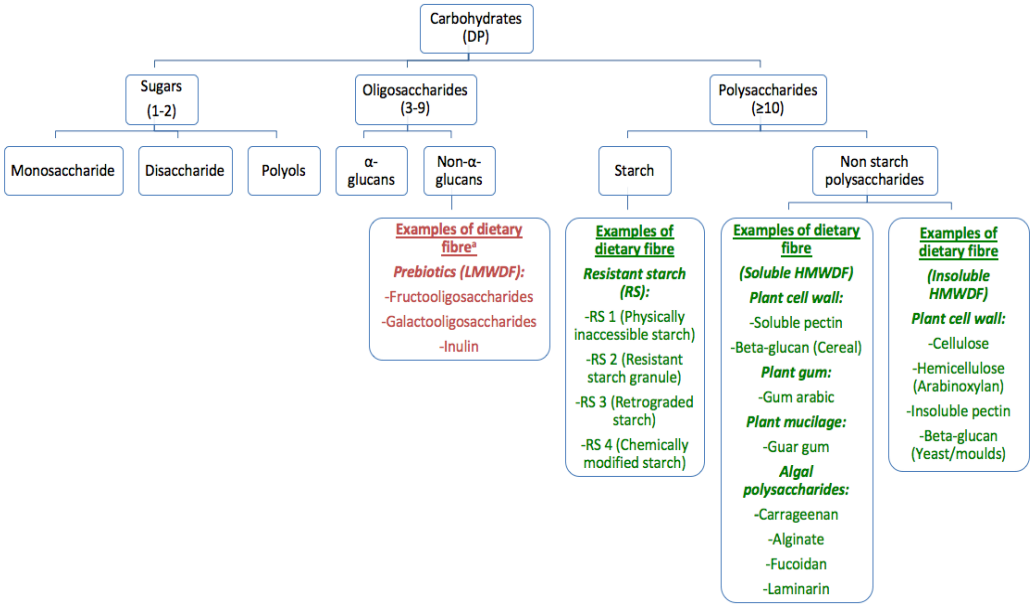


Figure 2.2: Classification of carbohydrates by degree of polymerization.

Information sourced from: Cummings and Mann (2012); Gibson et al. (2010); Westenbrink et al. (2013). Abbreviations: DP, degree of polymerization, also known as the number of monomeric, single-sugar units; LMWDF, low molecular weight fibre; HMWDF, high molecular weight fibre; RS, resistant starch. Examples of fibre are labelled in green. [a] Prebiotics are labelled in red.

Prebiotics are classified as carbohydrates and may also be classed as fibre (see non-α-glucans in **Figure 2.2**). Prebiotics are defined as 'selectively fermented ingredients that result in specific changes in the composition and/or activity of the gastrointestinal microbiota, thus conferring benefit(s) upon host health' (Gibson et al., 2010). To be classed as a prebiotic, a substance has to, therefore (Gibson et al., 2010):

1. Be resistant to gastric acidity and hydrolysis by mammalian enzymes and gastrointestinal absorption.
2. Be fermented by intestinal microbiota.
3. Be able to selectively stimulate the growth and/or activity of intestinal bacteria associated with health and wellbeing.

Therefore, when comparing dietary intakes from different gut microbiota studies, it is important to consider the source of the food composition values which food composition database was used and what chemical analytical method was used. It would also be good practice if attempting to replicate a study, to use the nutrient values from the same database.

2.4 FFQ Validation

2.4.1 Study design

As mentioned in section 2.3.2, an FFQ needs to be validated in the population it is used in for the researcher to determine how true a representation it provides of what has been consumed. As absolute validity is hard to determine practically, validity relative to another dietary assessment method is usually assessed instead. The main considerations for design of an FFQ validation study are summarised in **Table 2.4**.

Table 2.4: Summary of considerations for design of an FFQ validation study.

Factor	Comments
Reference method	A dietary method that has independent errors from the errors associated with FFQs – particularly memory and estimation errors. Biomarkers that provide an independent assessment of the intake of a particular nutrient, and respond in a dose-dependent manner (e.g., sodium from 24-hour urinary sodium) can also be used for some nutrients.
Study population	FFQs should be validated with participants who are recruited from the population of interest. The participants' ethnicity, weight status, and dieting history should be taken into consideration where relevant. A sample size of at least 100 participants is generally required.
Age of participant	The validation study should consider memory (because of recall) and conceptualization (because of the need to estimate portion sizes) skills that affect the responses of both younger and older participants. In certain cases, primary caregivers may be asked to act as proxies for completing the questionnaires.
Nutrient(s) of interest	FFQs should be designed to capture the nutrient(s) of interest if this is the intended purpose. For example, having a food list that captures most of the possible foods to the level that is appropriate. Reference dietary methods should also be able to measure the nutrient(s) of interest.
Sequence of administration	The problem of administering the FFQ *before* the reference dietary method is that it is assessing a different time frame from the reference dietary method. On the other hand, the problem of administering the FFQ *after* the reference dietary method is that the FFQ may be completed better than usual as the participants have a better understanding of dietary assessment. One option is to take a random first or second FFQ so that the effects of both are included.
Time frame	The reference dietary method should measure the same time frame as the FFQ. The FFQ should be administered at least twice if reproducibility is to be measured.

Information from Cade et al. (2002); Gibson (2005); Willett (1998).

2.4.2 Statistical analysis

The statistical methods used to analyse the relative validity of an FFQ depend on the objective of the study, for example, is the objective to determine group mean intakes, or to rank intakes? Generally, several different statistical methods are used, and the results are compared, but need to be interpreted with caution. For example, Bland and Altman plots (Bland & Altman, 1999) are commonly reported even though they show wide limits of agreement for nutrient intakes determined by FFQs, confirming that FFQs are not useful for describing the nutrient intakes of individuals. In addition, correlations between nutrients collected via the FFQ and reference method are reported in virtually all FFQ validation studies. Although there are a number of well-accepted

problems with this statistical approach (Bland & Altman, 1999), correlation coefficients are useful for indicating whether the FFQ is useful for ranking nutrient intakes.

Each nutrient should be considered separately, as an FFQ may be valid for some but not all nutrients. The statistical methods most commonly used to assess relative validity are summarized in **Table 2.5**.

Statistical methods such as comparison of mean nutrient intakes and correlations between nutrients estimated via the test (FFQ) and reference dietary method, can be used to assess the FFQ's validity, and reproducibility can be assessed by comparing the results of a first and second administration of the FFQ as described in **Table 2.5**.

Table 2.5: Summary of common statistical methods used in FFQ validation.

Statistical method	Comments
Comparison of means – t-test	If the nutrient data are normally distributed, **paired t-tests** can be used to examine whether the group means for each nutrient of interest estimated by the FFQ and the reference method are different.
Comparison of means – Wilcoxon's signed rank test	If the nutrient data are skewed, and not suitable for log-transformation, the **Wilcoxon's signed rank test** for paired data can be used to test whether the group median intakes estimated by the FFQ and the reference method differ significantly.
Correlation – Pearson's	**Pearson's correlation** can be used to measure the strength of the association at the individual level between intakes estimated by the FFQ and the reference method. This provides a measure of how well the FFQ ranks intakes. If the nutrient data are skewed, the data should be transformed before calculating the correlation coefficients.
Correlation – Spearman's	**Spearman's correlation** can be used to measure how well the FFQ ranks intakes when data are skewed and not suitable for log-transformation.
Correlation – intraclass	**Intraclass correlation** can be used to measure the strength of the association at the group level between intakes estimated by the FFQ and the reference method. It considers the extent of disagreement within pairs and the degree of correlation.
Deattenuated correlation coefficients	Nutrient intakes differ within one participant over time, which is known as intra-individual variation. This may affect the correlation coefficients between intakes estimated by the FFQ and reference method. **Deattenuated** nutrient intakes for the reference method can be used in the calculation of correlation coefficients to address this. This can be done using the Multiple Source Method (MSM) program (Harttig, Haubrock, Knüppel, & Boeing, 2011).
Interpretation of correlation coefficients	***Poor association***: <0.20 (Masson et al., 2003) ***Acceptable association***: 0.20 – 0.49 (Masson et al., 2003); 0.30 – 0.49 (Willett, 1998) ***Good association***: ≥0.50 (Masson et al., 2003); 0.50 – 0.70 (Willett, 1998)

Statistical method	Comments
	Care needs to be taken in the interpretation of correlation coefficients as a high correlation coefficient does not necessarily mean that there is good agreement between the test and reference method (e.g., if the FFQ consistently estimates values 30% greater than the reference method, correlation would be high but the agreement is unsatisfactory). These descriptors for correlation coefficients should, therefore, be used as indicators of the ability of the FFQ to rank intakes, not to determine absolute intakes.
Cross-classification	In **cross-classification**, participants are classified into thirds (at tertiles), fourths (at quartiles), or fifths (at quintiles) of intake by the FFQ and reference method. Usually, the percentage of participants correctly classified into the same category, and grossly misclassified into the opposite category is calculated. One of the disadvantages of cross-classification is that the percentage agreement will include the agreement by chance, and the degree of the percentage by chance is dependent on the number of categories used (thirds, fourths, or fifths).
Kappa coefficient	Similar to cross-classification, **Cohen's kappa coefficient** is a statistic that not only measures inter-rater agreement for categorical items, but also considers the possibility of the agreement occurring by chance. One of the disadvantages of the kappa coefficient is that it depends on the number of categories used and also the weightings applied to the categories.
Absolute values for surrogate categories	In the **absolute values for surrogate categories** approach, participants are assigned to ranked groups (by quartiles or quintiles) according to intake estimated by the FFQ, then the mean intake in each group is calculated using the intake determined by the reference method. Regression models are fitted to see whether there is a trend in the step-wise increases across the groups, and the difference in the extreme categories is calculated.
Bland and Altman	The mean and standard deviation of the difference between the FFQ and reference method is calculated for each nutrient of interest. The 95% limits of agreement also plotted, which is the mean difference ± 1.96 standard deviation of the differences, provide an interval within which 95% of differences between measurements by the FFQ and reference method are expected to lie. **Bland and Altman plots** can then be plotted using the mean intake from both FFQ and reference method against the difference of the two methods. These plots give a visual assessment of the agreement across all intake levels, show outliers, illustrate the range of agreement, and demonstrate patterns of bias in the data for the nutrient of interest.

Information from Gibson (2005); Lombard, Steyn, Charlton, and Senekal (2015); Willett (1998).

Overall, when conducting an FFQ validation study, a combination of statistical analyses should be used to get a better overview of the performance of the FFQ for estimating intake of the nutrients of interest.

2.5 Introduction to the gut microbiota

The term 'microbiome' refers to the entire habitat, consisting of all the microorganisms (archaea, bacteria and viruses) (Marchesi & Ravel, 2015), their genes, and the surrounding environmental conditions, but most discussions in the health literature use the term to refer to the bacterial community. The colon contains a microbial community (the gut microbiota), composed mostly of obligately anaerobic bacterial species, which digest the indigestible components of the diet (fibre fractions such as cellulose, hemicelluloses, resistant starch) consumed by the human host (Tannock, 2017). In adult human large intestines, the dominant species are from the families Bacteroidaceae, Prevotellaceae, Rikenellaceae, (Bacteroidetes phylum), and Lachnospiraceae and Ruminococcaceae (Firmicutes phylum) (Donaldson, Lee, & Mazmanian, 2016); while in infants, the dominant genera are *Bifidobacterium* (Actinobacteria phylum)*, Veillonella, Streptococcus, Clostridium* (Firmicutes phylum), *Citrobacter, Escherichia,* (Proteobacteria phylum), and *Bacteroides* (Bacteroidetes phylum) (Milani et al., 2017). To date, even though around 1500 bacterial species have been identified that may be able to inhabit the human colon (Zou et al., 2019), it appears that each adult human has a microbiota containing between 160 to 400 species (Lloyd-Price, Abu-Ali, & Huttenhower, 2016). Moreover, the microbiota does not have the same composition in every human (microbiotas are individualistic) although broad patterns can be recognized at higher taxonomic levels.

In the colon, dietary fibre is hydrolysed and fermented by the combined activities of the microbiota to produce mainly gases (hydrogen, carbon dioxide, methane) and short chain fatty acids (SCFAs; principally acetate, propionate, butyrate; present as an approximate molar ratio of 60:20:20) (Koh, De Vadder, Kovatcheva-Datchary, & Bäckhed, 2016) (den Besten et al., 2013). The SCFAs are absorbed from the colon and provide energy, about 10% of daily caloric requirements (Bergman, 1990), so the microbiota has primary importance to humans in harvesting extra energy from the diet that could not be obtained if colonic bacteria were absent (Schwiertz et al., 2010). Moreover, increasing numbers of studies indicate a role for SCFAs in maintaining the host's metabolic stability (Makki, Deehan, Walter, & Bäckhed, 2018). In addition, the microbiota provides benefits to the host by 'colonization resistance', where pathogens are inhibited by other bacteria so that they do not colonize the gut (Donaldson et al., 2016).

In contrast, alterations or imbalances in microbiota composition (dysbiosis) have been associated with a wide range of health conditions including: obesity, diabetes, inflammatory bowel disease, irritable bowel syndrome, and allergies (Bäckhed et al., 2005; Flint, Scott, Louis, & Duncan, 2012b; Honda & Littman, 2012; Ringel & Carroll, 2009; Robles Alonso & Guarner, 2013).

2.6 Development of the gut microbiota

The newborn gut at birth has an aerobic environment which supports the growth of facultative anaerobes such as members of the Enterobacteriaceae family (Arrieta, Stiemsma, Amenyogbe, Brown, & Finlay, 2014). Within a few days post birth, the environment turns anaerobic which supports colonization by obligate anaerobes such as *Bifidobacterium*, *Clostridium*, and *Bacteroides* (Matamoros, Gras-Leguen, Le Vacon, Potel, & de La Cochetiere, 2013). Depending on the mode of delivery, for the first few weeks of life the infant gut resembles either the microbiota of the vagina (vaginal delivery) or maternal skin (caesarean delivery) (Dominguez-Bello et al., 2010). As the infant's energy source is mainly milk for the first 4-6 months, milk lactose and oligosaccharide fermenters such as *Bifidobacterium* thrive (Turroni et al., 2012). In particular, a uniqueness of human milk is the high abundance of human milk oligosaccharides (HMO) (13 – 21 g/L), consisting of over 130 different types. In human milk, lactose is the main carbohydrate (56 – 69 g/L at 4 months) and monosaccharides represent 1.2% of the total carbohydrates (Coppa et al., 1993). HMO are composed of five monosaccharides, glucose, galactose, *N*-acetylglucosamine, fucose and sialic acid (Bode, 2012). Infant formula has oligosaccharides added to it. However, the oligosaccharides in infant formula are still found to be structurally different from the oligosaccharides naturally occurring in human milk (Bode & Jantscher-Krenn, 2012).

After the introduction of solids (i.e. during complementary feeding), the infant's gut microbiota drastically changes, and it becomes more adult-like at around 2 to 5 years (Fallani et al., 2011; Koenig et al., 2011; Palmer, Bik, DiGiulio, Relman, & Brown, 2007), although one study has still found gut instability as late as pre-adolescence (Hollister et al., 2015). In general, the infant gut microbiota is characterized by low diversity and a greater relative abundance of the Proteobacteria and Actinobacteria phyla, while the adult microbiota has high diversity with dominance of Firmicutes and Bacteroidetes phyla (Bäckhed, 2011). However, there is still very

limited information about how a pre-school child's diet affects their gut microbiota in this period of rapid growth (Laursen et al., 2017).

2.7 Measurement of the gut microbiota

At the simplest level, questions about the gut microbiota can be stated as: 'Who is in the bacterial community of these participants?', 'What is the relative abundance of each type of bacteria?', 'What functions could these bacteria carry out?', and 'What functions are they actually carrying out?'. Key methods for answering these questions can be found in **Figure 2.3**.

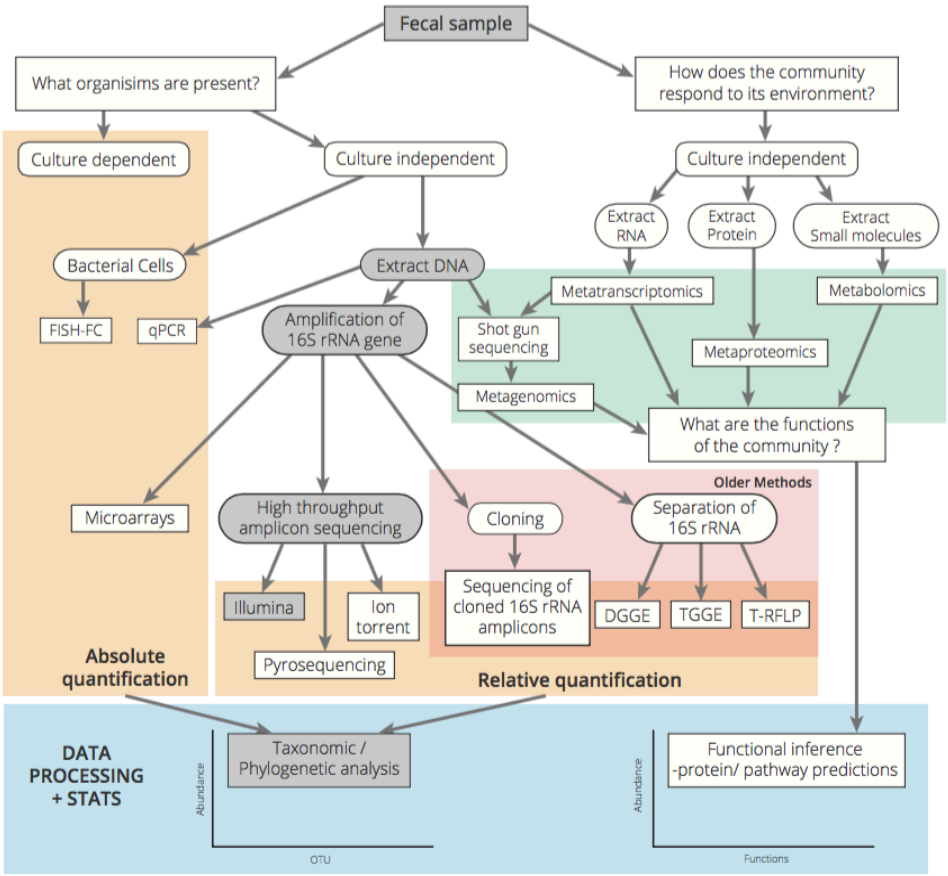


Figure 2.3: Overview of some of the common techniques used in the study of human gut microbiota.

The grey highlighting shows the 16S rRNA pathway in relation to the many other pathways that can be taken. Carrying out the methods under the 'What organisms are present?' pathway would allow us to determine 'Who is in the bacterial community of these participants?', and 'What is the relative abundance of each type of bacteria?'. Carrying out the methods under 'How does the community respond to its environment?' would allow us to determine 'What functions could these bacteria carry out?', and 'What functions are they actually carrying out?'. The Candidate used SmartDraw software for the initial design of this figure and acknowledges Kylie Paterson as the graphic designer for this figure, with content provided by the Candidate.

In the past, microbiologists relied on traditional methods such as culture (growing bacteria in the laboratory), and various methods of identification of the bacteria that were cultured (including microscopy to determine cell shapes) to determine the bacterial species, and their abundance, in faecal samples (Tannock, 2005). However, the traditional culture dependent methods are limited because many bacteria have not yet been grown, and because the vast number of bacteria in the human gut means the process is extremely time consuming and logistically difficult. Detailed explanations of culturing techniques are available in the literature (Lagier et al., 2015a; Lagier et al., 2015b). With the introduction of high throughput sequencing (HTS) in 2005, also known as 'culture independent methods' (McPherson, 2014; Walker, 2016), many more types of bacteria can be identified in a sample and it is possible to determine how abundant they are much more quickly.

The microbiota data in this thesis were obtained from HTS of the 16S rRNA gene. Hence, this review focuses on the use of the 16S rRNA gene to identify and quantify the bacterial taxa in the gut.

2.7.1 High throughput sequencing of 16S rRNA gene

A typical diet and gut microbiota study describing the presence and abundance of different bacterial taxa follows a sequence of activities (**Figure 2.3**, pathway highlighted in grey) with 4 key gut microbiota analysis steps:

a. Collection of faecal samples

Faeces are generally used to determine the composition of the colonic microbiota. Although faeces represent the end of the digestive process and are far away from the site of maximum microbiota activity (the proximal colon), faecal samples can be obtained non-invasively and provide an indication of the types of bacteria present in the colon and the proportions (relative abundances) in which they occur (at least in the rectum). Moreover, faecal samples are less invasive than a biopsy, are relatively fast to collect, and have less respondent burden for participants (Aguirre & Venema, 2015).

Culture independent methods are preferred because, unlike culture dependent methods which require fresh faeces, faeces can be frozen or chemically preserved after collection, so that studies with numerous participants, in diverse locations, are logistically feasible. However, culture independent methods do have some limitations.

For instance, some studies have found systematic error with a decrease in *Bacteroides* spp. after one week of storage at -20°C (Maukonen, Simões, & Saarela, 2012), and an increase in the 'Firmicutes to Bacteroidetes' (F/B) ratio in frozen samples compared to fresh samples (Bahl, Bergström, & Licht, 2012). In addition, ice formation during freezing may cause damage to the bacterial cells, and hence cryoprotectants such as glycerol may be added to prevent the breakage of bacterial cells caused by the ice crystals (Bircher, Schwab, Geirnaert, & Lacroix, 2018). Therefore, for best practice, it is recommended that samples are stored for no longer than 15 minutes at room temperature, and for no more than three days in a domestic -20°C freezer (Gorzelak et al., 2015), and if using a cryoprotectant for storage, it is important to ensure that all the samples are stored in the same manner (Pollock et al., 2018). After this, the samples are usually transported to be stored in the laboratory in a -80°C freezer before further analysis. In addition, because bacteria are present everywhere in our environment, it is good practice to ask participants to wear sterile gloves while collecting the faecal sample to prevent contamination of the sample. In studies looking at children, faeces are usually obtained from a nappy (diaper) if the child is still using them, or with the aid of the caregiver when the child is using the toilet.

b. Extraction of DNA

The next step is to extract the deoxyribonucleic acid (DNA) from the faecal samples. There are Gram-positive and Gram-negative bacteria. Gram-positive bacteria cell walls are thicker than Gram-negative bacteria (Olsen, Lane, Giovannoni, Pace, & Stahl, 1986). Hence, to ensure representative proportions of the bacteria from the faecal sample, it is important that mechanical lysis is used in addition to chemical methods, so that the nucleic acid is obtained even from cells that are difficult to break open by chemical methods (de Boer et al., 2010; Lawley & Tannock, 2012). This is important in infants because, if the detection of bifidobacterial (the main bacteria in infants; Gram-positive) is not optimised (i.e. using mechanical methods), this would lead to an erroneous conclusion of the absence of bifidobacterial in the following steps (Walker et al., 2015).

c. Amplification of the 16S rRNA gene

Direct sequencing of 16S rRNA amplicons (piece of DNA that contains replicated genetic material) is commonly used to determine the types and abundance of

bacteria present in a faecal sample (**Figure 2.3**). The 16S rRNA gene sequence of bacteria is used because it contains highly conserved regions of nucleotide base sequence across the bacterial world (Janda & Abbott, 2007; Woese, 1987). In addition, these conserved regions are interspersed with variable regions (V regions) that contain 'signatures' of phylogenetic groups (Tannock, 2017).

In order to have enough DNA for sequencing, the 16S rRNA gene needs to be amplified (i.e. many copies need to be made) by Polymerase Chain Reaction (PCR). Conserved regions of the 16S rRNA gene sequence are used as molecular targets that primers (short sequences of specific base pairs) can attach to so that the PCR can identify the specific section of DNA that is of interest. PCR is then run to produce amplicons of the V regions.

There are a few issues with the use of PCR. First, the primers may not be optimised to detect certain genes and this may lead to an underrepresentation of certain bacteria (Walker et al., 2015). Second, chimeras (amplicons that are generated as an artefact of the PCR process, so do not reflect DNA sequences that were in the original faecal sample) may be formed (Walker, 2016). For example, a chimera can be formed when polymerases do not fully copy the whole strand. In a subsequent cycle, part of the incomplete amplicon may be used incorrectly as a primer with the resultant template producing aberrant amplicons that do not reflect the genetic material that was in the original faecal sample. To address the problem of chimera formation, chimera checking is performed by software such as vsearch, where chimeras are detected either *de novo* or with a reference database using the UCHIME algorithm (Rognes, Flouri, Nichols, Quince, & Mahé, 2016).

d. Sequencing

Current analyses mostly use HTS methodology provided by the Illumina platform which generates millions of short sequences from each sample that can be assembled and aligned with taxonomic databases to provide an overview of the bacterial taxa in the sample and their relative abundances.

Adaptors (20-30 base pairs) are attached to the primers (anchors) described above. These adaptors cause the amplicons to bind to a plate so that they can be sequenced. A range of methods used for HTS (also called 'next generation sequencing'), namely pyrosequencing, Illumina and Ion Torrent, have been

summarized in the literature (Kuczynski et al., 2012; Lawley & Tannock, 2012; Lay, 2009; Walker, 2016).

Multiplexing of samples can be used to make this sequencing more cost effective. In effect, a unique identifier (DNA code) is attached to the primers for the DNA from each faecal sample. This means that the DNAs from multiple faecal samples can be combined and analysed in a single run, and then the data can be assigned back to the correct faecal sample by separating the reads, using the unique identifiers.

The sequenced reads then need to be identified. First, the reads are clustered by 97% similarity; that is, algorithms are used to group the sequences based on 97% sequence similarity. There are different algorithms (e.g., Uclust, Mothur, CD-HIT, etc.) that are used but this adds another layer of complexity and makes different studies in the literature harder to compare (Chen, Zhang, Cheng, Zhang, & Zhao, 2013). It is often assumed that this process clusters sequences at a species level, but the clusters are more correctly called 'operational taxonomic unit' (OTUs) (Konstantinidis & Tiedje, 2005). These reads are then assigned to their different taxa by comparing the sequences with several large databases of reference sequences and taxonomies, such as Greengenes (DeSantis et al., 2006), SILVA (Pruesse et al., 2007) and the Ribosomal Database Project (Cole et al., 2009). It is important for researchers to describe the methods used at each step of the process, so that comparison of the results from different studies can consider the different methods used.

2.7.2 Other methods for describing the gut microbiota

Even though HTS of 16S rRNA from faecal samples is a very powerful tool, it can only answer the questions 'Who is in the bacterial community of these participants?' and 'What is the relative abundance of each type of bacteria?' (**Figure 2.3**). To determine functional capacity ('What functions could these bacteria carry out?'), and actual activities ('What functions are they actually carrying out?'), other methods need to be used. The methods for determining functionality and actual activities of the gut microbiota were not carried out in this thesis; but for completeness, are explained briefly in the paragraphs below.

a. Determining the functional capacity of the microbiota

The functional capacity of the microbiota can be determined using a 'shotgun sequencing' metagenomics approach. In shotgun sequencing, bulk DNA is extracted from faeces and mechanically shredded into small pieces and these fragments are sequenced randomly (i.e. particular genes are not amplified) (Allaband et al., 2019). Computer software (such as MetaPhlAn, HUMAnN, MaAsLiN) is used to assemble the sequences into recognizable genes, which are then aligned with taxonomic databases to identify the origin of 16S rRNA gene sequences (phylogeny) and the gene categories that are present in relation to biochemical pathways (e.g., as provided by KEGG) (Kuczynski et al., 2012). By gathering counts of different categories of genes in these biochemical pathways, the metagenome (comprising the total genomic content of the microbiota) can be determined and compared between faecal samples. Metagenomic data shows the biochemical potential of the microbiota (what the microbiota has the genetic capacity to do). However, it does not tell us what the microbiota was actually doing at the time the faeces were collected.

b. Determining actual activities of the microbiota

Methods such as metatranscriptomics, metaproteomics and metabolomics can be used to determine different aspects of the actual activities of the microbiota. Metatranscriptomics is the study of the RNA transcripts of an entire microbial community so gives an insight into the functional activity of the microbiota – telling us which genes were transcribed at a particular moment in time – and therefore the functions the bacteria were carrying out at the time the sample was collected (Kuczynski et al., 2012). Metaproteomics is the study of the proteins produced by the microbiota – giving insight into how the microbiota was responding to its environmental conditions at the time the sample was collected (Wilmes & Bond, 2009). Metabolomics is the study of the metabolites, such as short chain fatty acids, present within the faecal sample at the time of sampling. The SCFAs reflect the fermentation of carbohydrates by the gut microbiota. Metaproteomics and metabolomics allow for direct monitoring of the end products of bacterial metabolism, which cannot be done via metatranscriptomics, metagenomics, or 16S rRNA gene sequencing (Ursell et al., 2014). It is important to decide early on in the design of the study what methods are going to be used to analyse the gut microbiota because different collection and storage

methods are needed for different analyses. For example, for metatranscriptomics, faeces must be stored in a solution that prevents the destruction of mRNA because it rapidly degrades after synthesis (it has a short half-life) (Reck et al., 2015). Chemical solutions such as RNALater (Reck et al., 2015) and polyvinyl sulfonic acid (PVSA) (Earl, Smith, Lease, & Bundy, 2018) provide ribonuclease inactivation and hence are useful chemical protectants to prevent mRNA destruction.

2.8 Gut microbiota data and their statistical analysis

Gut microbiota data generated from the 16S rRNA sequencing steps mentioned above (section 2.7.1) generate data (i.e. OTU counts, taxa relative abundance) with a number of challenging features, as described below.

2.8.1 The nature of gut microbiota data

a. Microbiota data are compositional

Comparisons of microbiota data, in simple terms, involve counting the number of genes indicative of a particular taxon and normalizing the information in terms of percentage of the total number of gene sequences in the sample. This is known as the 'relative abundance'. Relative abundances are compositional data because the sum of all relative abundances of the different bacterial taxa in a sample is constrained to 100%. This introduces a co-dependency between the taxa so that if the percentage of one bacterial taxon increases, then the percentage of at least one other bacterial taxon has to decrease because by definition the sum must always be 100% (Gloor, Macklaim, Pawlowsky-Glahn, & Egozcue, 2017). In addition, the actual total sum of all the OTUs is arbitrary and dependent on the sequencing instrument used – hence the need to use 'relative' abundance (Xia et al., 2018). It is important therefore that statistical analyses take into account the compositional nature of the data and perform appropriate transformations of the data before using traditional statistical methods (Aitchison & Egozcue, 2005).

b. Microbiota data are high dimensional and underdetermined

Using HTS methods, in typical microbiome studies, the number of taxa or OTUs reported may number in the hundreds, or even thousands. This is described as 'high dimensional' data as each bacterial taxon represents one dimension, hence, when looking at the relationships of the hundreds of bacteria, there will be hundreds of

dimensions (Li, 2015; Shankar et al., 2015). Microbiota data are also 'underdetermined' because the number of taxa or OTUs (which is usually in the hundreds) is much greater than the number of faecal samples or participants (which is in the tens in many studies) (Xia et al., 2018). These characteristics of the data are challenging when it comes to statistical analysis because it results in what is known as a large *p* and small *n* problem, where in simple terms the *p* represents the 'unknown' variables (or bacterial taxa), and *n* represents the samples (or participants) (Johnstone & Titterington, 2009). In the classic text by Huber (1981) it was recommended that '$n/p \geq 5$' is a plausible rule of thumb for good data-analytical practice.

c. Microbiota data are often sparse with many zeros

Microbiota data can be described as 'sparse' because many taxa will be absent in any one sample. This means that the data can contain many zeros. This leads to (Xia & Sun, 2017) data with a truncated normal distribution, and can also cause problems for some statistical procedures such as log-transformation (which cannot be applied to zero values) or ratios (where the denominator cannot be zero).

d. Microbiota data are highly complex

Moreover, microbiota data are extremely complex, with multiple variables that interact with each other. Examining each separate taxon is not very informative without considering the rest of the taxa. This is problematic for research studies investigating the relationship between characteristics of the diet and the gut microbiota, or between the gut microbiota and health where the gut microbiota would ideally be described using summary variables that capture specific communities present in the gut or other useful constructs. Gut microbiota data need to be summarised in some way to be useable. Bacterial taxonomy, alpha and beta diversity, and clusters are three ways that the microbiota data are commonly examined in diet and microbiota studies (section 2.8.2).

2.8.2 Methods for summarising gut microbiota data

a. Description of the bacterial taxa present

Taxonomy is defined as the science of biological classification (Willey, Sherwood, & Woolverton, 2013). There is a hierarchical arrangement in taxonomy, starting with the Kingdom (the American system has six kingdoms: Animalia, Plantae,

Fungi, Protista, Archaea, Bacteria), Phylum (some common phyla in the infant gut are: Actinobacteria, Proteobacteria, Bacteroidetes, Firmicutes, Verrucomicrobia) (Bäckhed et al., 2015), Class, Order, Family, Genus (the generic name; indicated with a capital letter and italics) and Species (the specific name; indicated by italics). Microorganisms are named using the binomial system of Linnaeus, which consists of a Latinized, italicized name consisting of two parts. The first part is the generic name (i.e. the genus), and the second is the species name. For instance, *Faecalibacterium prausnitzii*, the generic name is *Faecalibacterium* and the specific name is *prausnitzii* (this should not be written on its own) (**Figure 2.4**). The specific name is stable but the generic name can change if the microorganism is assigned to another genus. The term 'taxon' (plural 'taxa') refers to each level of the bacterial hierarchy in general, that is, 'species' is a bacterial taxon, and 'genus' is also a bacterial taxon.

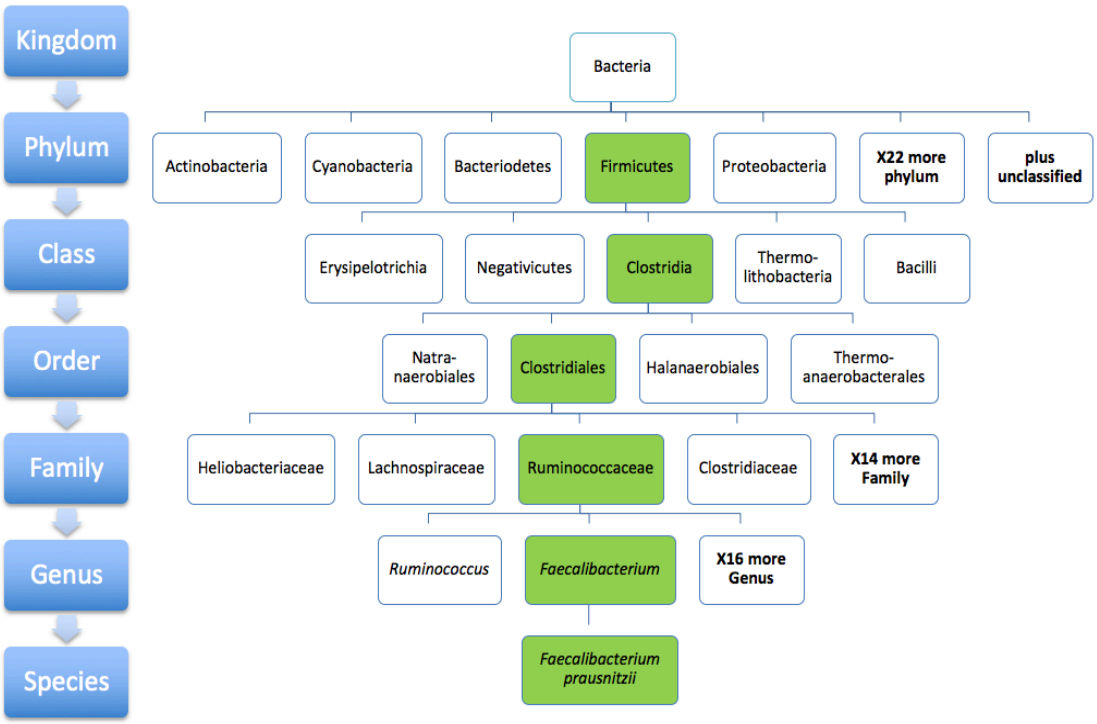


Figure 2.4: An example of the hierarchical arrangement in taxonomy of *F.prausnitzii*.

The **green highlight** shows the taxonomic lineage of *F.prausnitzii*. *F.prausnitzii* is found to be abundant in healthy adult gut and diminished in diseased states such as Crohn's disease (Flint et al., 2012b).

The best-case scenario would be to examine the gut microbiota taxa at the species level, as this would give us the most information and enable applications in medicine

and the food industry. However, examining a single species without consideration of all other species may give rise to misleading results, given the interaction and co-dependence of microbiota data. Furthermore, as mentioned in section 2.7.1, the sequences generated are clustered into OTUs at a chosen similarity level to get different taxonomic data. Metagenomic sequencing to get whole 16S rRNA gene sequences and other sequences in the genome that are taxonomically useful, is required to get species-level information. 16S rRNA sequencing analyse partial gene sequences, hence the precision for identification is not as great. Generally, when using 16S gene data, researchers are more confident to identify family-level taxa. However, genus-level taxa are still acceptable to be identified using 16S gene data (Li, 2015; Oono, 2017).

b. Alpha and beta diversity

Two diversity indices are commonly calculated. First, 'alpha diversity' looks at the variety of bacteria present. This can be quantified by the number of observed species (a measure of alpha diversity richness, i.e. number of different species in a sample), and by the Shannon (Shannon, 1948) or Simpson (Simpson, 1949) indices (measures of alpha diversity richness and evenness i.e. the number of different species and how balanced the distribution of the different species is in a sample).

Second, 'beta diversity', which is also known as an index of similarity, is a description of how many taxa are shared between two communities. This can be quantified by Bray-Curtis dissimilarity (Bray & Curtis, 1957), where the higher the beta diversity, the less similar the communities. Beta diversity also has a phylogenetic aspect (quantified by UniFrac, weighted and unweighted), which shows how different OTUs diverged from each other in evolutionary terms (Lozupone & Knight, 2005; Lozupone, Lladser, Knights, Stombaugh, & Knight, 2011). These analyses are carried out using software pipelines (such as QIIME (Kuczynski et al., 2011)) that are freely available through the internet.

c. Clustering

Principal Co-ordinate Analysis (PCoA) is one type of clustering method that uses dissimilarity (Bray-Curtis) or distance (UniFrac) (Gower, 2005). It is similar to the concept of Principal Component Analysis (PCA) which uses dimension reduction. PCoA provides information on visualisation of the dissimilarities of the data (Gower, 2005).

Another clustering method is enterotyping which was introduced in a seminal paper in 2011 (Arumugam et al., 2011) and has mainly been used to describe adult gut microbiota. Enterotyping is a way of categorising microbiotas based on recognizing clusters of OTUs/ species and assigns an individual to a cluster (called an enterotype). The original study described a *Bacteroides* predominant, a *Prevotella* predominant, and a *Ruminococcus* predominant enterotype (Arumugam et al., 2011). Enterotypes have been used in research investigating diet and obesity in adults (Christensen, Roager, Astrup, & Hjorth, 2018; Wu et al., 2011), and recently in one study in children (Zhong et al., 2019). One of the limitations of enterotying is that because gut microbiota data are continuous and vary widely within individuals (Costello et al., 2009), assigning each participant to a single enterotype might over-simplify such complex data (Jeffery, Claesson, O'Toole, & Shanahan, 2012; Knights et al., 2014). Moreover, a recent perspective written, which included the authors of the seminal paper on enterotypes, acknowledged that defining meaningful and robust boundaries are a challenge, and that the enterotyping method may not be a statistically rigorous approach (Costea et al., 2018).

2.8.3 Statistical analysis of gut microbiota data

A wide range of software programs are used by bioinformaticians to carry out statistical analysis of gut microbiota data. These include bioinformatic software such as STAMP (which provides a user-friendly graphical interface for generating exploratory plots; and implements statistical tests such as ANOVA, Kruskal-Wallis, PCA and presents confidence intervals (Parks et al., 2014)), and MicrobiomeAnalyst (which is a web-based tool which provides similar functions to STAMP and also includes comparative analysis and taxon set enrichment analysis (Dhariwal et al., 2017)).

Researchers in the nutrition and health research field use biostatistics which encompasses a range of statistical techniques applied to answer pre-specified research questions. A number of commonly used biostatistical methods are also appropriate for analysing gut microbiota data. An advantage of using these methods rather than using programs that analyse the gut microbiota data for you, is that biostatistical analyses are not dependent on one particular statistical analysis programme or package, because they are stand-alone methods. Furthermore, the methods can be tailored to answer the research question at hand and their performance, assumptions and errors can be more

readily identified and addressed. **Table 2.6** gives a summary of some of the key biostatistical methods that have been used in gut microbiota analyses.

For many of the statistics used, false discovery rate (FDR) adjustments are applied to p-values due to the multiple comparisons made in microbiota studies. This includes Bonferroni correction (Simes, 1986) and the Benjamini-Hochberg (BH) procedure (Benjamini & Hochberg, 1995). For improving reproducibility of the data, it is important to state what level of FDR is being used in the analysis.

For the studies reported in this thesis (**Chapters 4, 5, and 7**), the emphasis is on carrying out good statistical practice, where the results are to our best effort, interpreted in context and with complete reporting and explanation of what the various data summaries mean, and not based on p-value cut-offs. As stated by the American Statistical Association (ASA), on p-values, 'no single index (e.g., p-value) should substitute for scientific reasoning' (Wasserstein & Lazar, 2016).

Table 2.6: Summary of key biostatistical methods that have been used in gut microbiota analyses.

Statistical method	Comments
Comparison of means – 2-sided t-test	**T-tests** can be used to examine whether the means are different between groups or within pairs.
Correlation – Spearman's	As relative abundance data are usually non-normally distributed, **Spearman's correlation** is used instead of Pearson's correlation, to measure the strength and direction between two ranked variables (e.g., relative abundance of a bacterial taxon and intake of a dietary component).
Regression models - Linear	In **regressions**, other variables (or confounders) can be added to the model to determine their effect on the regression estimate. Random effects can be used in a mixed effects regression model to account for correlated data (e.g. longitudinal data or clustered data). Effect estimates and confidence intervals are calculated for a broader understanding of associations, beyond just p-values.
Regression models - Logistic	Regression models can be built with a link function to suit the distribution of the outcome variable (e.g. a **logit** link function for binary data). It generates odd ratios.
Comparison of means – Wilcoxon rank sum test (also called Mann-Whitney U test)	If the data are skewed (non-normal), the **Wilcoxon rank sum test** can be used to examine whether the medians are different at the group level.
Comparison of means – Kruskal-Wallis test	**Kruskal-Wallis test** is the non-parametric (for non-normal data) equivalent to the one-way ANOVA. It extends the Wilcoxon rank sum test as it can test more than 2 groups.
Comparison of means – One-way ANOVA	**One-way ANOVA** generalizes the t-test to more than 2 groups. One-way ANOVA only has a single dependent variable.
Comparison of means – Multivariate ANOVA (MANOVA)	**MANOVA** is used when there are multiple dependent variables across the multiple groups that are of interest. The aim is to look for differences among groups in all dependent variables.
Comparison of means – Repeated measures ANOVA	**Repeated measures ANOVA** differs from MANOVA in that there is a single dependent variable, with multiple measurements for each participant.
Comparison of means – PERMANOVA	MANOVA assumes that the dependent variables are normally distributed within groups and hence have linear relationships. However, most gut microbiota data such as relative abundances are non-normal. Hence **PERMANOVA** is used because it does not require normally distributed data.

Information from Xia et al. (2018). Highlighted in grey are the preferred statistics for parametric analyses. Non-parametric and ANOVA statistics are not preferred as non-parametric analyses rely on p-value alone, rather than providing an effect size and a description of the data; and ANOVA is not good practice as the overall null hypothesis is that all the data come from groups that have identical means, and this is usually not the case for focussed research questions.

2.9 Factors that affect the gut microbiota

The main focus of this thesis is looking at the effects of diet on the gut microbiota. However, it is important to consider other factors that may affect the gut microbiota in children, as they can be confounders or interactions in the statistical models used to examine the effects of diet on the gut microbiota. The Candidate has referred to systematic and narrative reviews instead of individual studies for the factors that have been well studied. **Table 2.7** summarises a wide range of studies on factors that may influence the gut microbiota, but the list is not exhaustive. Please note – in this section, colour codes are used to differentiate different bacterial taxa as follows: Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria.

Pre-pregnancy body mass index (BMI) is one of the factors that affects the infant gut microbiota. In general, a straightforward association between pre-pregnancy BMI and alpha diversity has not been seen. However, when taking into account possible confounders, vaginally delivered infants with obese mothers have greater alpha diversity than those with normal weight mothers (Singh et al., 2019); infants with obese mothers who were not exclusively breastfed had greater alpha diversity than those with obese mothers who were exclusively breastfed (Sugino, Paneth, & Comstock, 2019). Moreover, vaginally delivered infants with obese mothers had higher *Staphylococcus* and *Enterococcus* (Singh et al., 2019), although this was in contradiction to another study that showed obese mothers have higher *Megasphaera* and lower *Staphylococcus* and *Streptococcus* (Sugino et al., 2019). These differences could be due to the different age of the participants (6 weeks vs 1 week) and the different confounders considered. In another study that considered confounders in the statistical models, pre-pregnancy BMI was not associated with bacterial taxa (Stanislawski et al., 2017).

Mode of delivery is one of the factors that has been well studied for its relationship with child gut microbiota. In general, caesarean delivered infants have lower alpha diversity at 4 months (Stinson, Payne, & Keelan, 2018). In addition, caesarean delivered infants were found to have lower abundance of *Bifidobacterium*, and *Bacteroides*, and higher abundance of *Clostridium* and *Lactobacillus* at 3 months (Rutayisire, Huang, Liu, & Tao, 2016). Similarly, caesarean delivered infants had higher abundance of *Clostridium spp.* at 1 year and lower abundance of *Escherichia–Shigella spp.* (Stinson et al., 2018). However, associations were found to have

disappeared after 3 months (Rutayisire et al., 2016), and after adjusting for other confounders (Stinson et al., 2018). Studies have been carried out which showed that the differences in gut microbiota associated with caesarean delivery may be partially restored by exclusive breastfeeding (Akagawa et al., 2019; Liu et al., 2019a), which may explain why associations between mode of delivery and gut microbiota composition were not seen after adjusting for other confounders such as infant feeding.

The effects of **infant feeding** on the gut microbiota are well-established. In a meta-analysis looking at 7 studies from 2014 to 2018, non-exclusively breastfed or shorter duration of exclusively breastfed infants had higher alpha diversity at 6 months (Ho et al., 2018). In addition, non-exclusively breastfed infants had a higher increase in abundance of *Bacteroides, Eubacterium, Veillonella* and *Megasphaera* than exclusively breastfed infants (Ho et al., 2018). One of the possible reasons exclusively breastfed infants have lower alpha diversity is because they have a predominance of Bifidobacteriaceae and *Bacteroides* while formula fed infants, although still having a predominance of Bifidobacteriaceae, had ~20% lower relative abundance of Bifidobacteriaceae (Tannock et al., 2013). This is supported by another study which found that formula fed infants had a wider spectrum of bacteria (e.g. *Bifidobacterium infantis*, *Clostridium perfringens*, *Clostridium difficile*, *Veillonella parvula* etc.) (Guaraldi & Salvatori, 2012). The high abundance of Bifidobacteriaceae and *Bacteroides* in exclusively breastfed infants can be related to the utilisation of HMO by species from Bifidobacteriaceae and *Bacteroides* (Marcobal et al., 2011). Moreover, the utilisation of HMO has been found to be strain-dependent in *in vitro* studies, with *B. infantis* degrading and utilizing intact HMO completely intracellularly, *Bifidobacterium bifidum* utilizing the glucose and galactose portions of HMO (Zabel et al., 2019), and *B. bifidum* acting as mediators for cross-feeding of HMO within bifidobacterial communities (Gotoh et al., 2018).

The effects of **antibiotics** on the gut microbiota have been widely studied and reviewed in both children and adults (Ferrer, Méndez-García, Rojo, Barbas, & Moya, 2017; Langdon, Crook, & Dantas, 2016). The use of antibiotics has been shown to be associated with decreased alpha diversity (Langdon et al., 2016). For example, in a randomised controlled trial (RCT), use of one of the commonly used paediatric antibiotics (Azithromycin) resulted in significantly lower alpha diversity (Simpson's Index) compared to the placebo (Oldenburg et al., 2018).

Parity is another factor that affects the child's gut microbiota. In general, having more siblings is associated with greater alpha diversity. Studies have seen different taxa associations with parity: negative associations with *Clostridium* and *Bacteroides* and a positive association with *Lactobacillus* in one study (Levin et al., 2016); and positive associations with *Haemophilus*, *Faecalibacterium*, *Barnesiella*, *Odoribacter*, *Asaccharobacter* and *Gordonibacter* in another (Laursen et al., 2015). As mentioned previously, the differences in associations with bacterial taxa may be due to the different timepoints considered (0 - 7 months vs 9 - 18 months). A possible explanation for the increased alpha diversity with having siblings may be the exposure to an environment with a greater variety of bacterial species at a time when the microbiota is still developing.

Table 2.7: Factors that affect the gut microbiota in children (the studies and reviews listed here are not exhaustive).

Factor	**(Reference)** **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Gut microbiota method**	**Main findings** **Confounding factor** 1. ***Alpha diversity, AD*** 2. ***Beta diversity/Clusters, BDC*** 3. ***Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. ***Other***	**Comments**
Pre-pregnancy BMI (or maternal weight status or gestational weight gain)	(Singh et al., 2019) USA New Hampshire Birth Cohort	6 wk 335 mother-infant pairs (male = 158)	Cross-sectional 16S rRNA, V4-V5 region, Illumina Mi-Seq	*Confounding factor:* mode of delivery *1. AD* (Observed OTUs, Shannon Index, Chao1): • Vaginal delivered infants born from obese mothers had greater AD than those born from normal weight mothers. • No differences in those born by caesarean. *3. Taxa* (Relative Abundance): <u>*Genus:*</u> • In vaginal delivered, infants born from obese mothers had higher *Staphylococcus* and *Enterococcus.* <u>*Species:*</u> • In vaginal delivered, infants born from obese mothers had overrepresentation of *Bacteroides fragilis*, *Escherichia coli* and *Veillonella dispar*.	• It is good to note that the study considered confounding factors that may affect the results.
Pre-pregnancy BMI (or maternal weight status or gestational weight gain)	(Sugino et al., 2019) USA ARCH$_{GUT}$ & BABY$_{GUT}$ cohort	1 wk 39 mother-infant pairs (male = 25)	Cross-sectional 16S rRNA, V4 region, Illumina Mi-Seq	*Confounding factor:* antibiotic exposure, EBF *1. AD* (Simpson Index, Shannon Index, Chao1): • No associations between AD and pre-pregnancy BMI. • After removing infants with antibiotic exposure, infants that were born from obese mothers and were EBF had lower AD than those non-EBF infants.	• The study removed infants with antibiotic exposure for certain analyses instead of using a statistical method to include those that were exposed to antibiotics. • Sample size is small for the multiple comparisons done.

Factor	(Reference) Country Study name if applicable	Participant age Sample size used in microbiota analyses (sex)	Study type Gut microbiota method	Main findings Confounding factor 1. *Alpha diversity, AD* 2. *Beta diversity/Clusters, BDC* 3. *Taxa; colour code:* Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. *Other*	Comments
				2. BDC (Bray-Curtis): • Pre-pregnancy BMI was not associated with infant gut community structure. *3. Taxa* (Relative Abundance): *Genus:* • Infants born from overweight or obese mothers had higher *Megasphaera* and lower *Staphylococcus* and *Streptococcus.*	
Pre-pregnancy BMI (or maternal weight status or gestational weight gain)	(Stanislawski et al., 2018) Norway Norwegian birth cohort	2 y 165 (male = 90) From 552 children	Cohort 16S rRNA, V4 region, Illumina Hi-Seq	*3. Taxa* (Relative Abundance): *Species:* • 60% (6/10) of the maternal species associated with maternal obesity was predictive of childhood BMI in the infant gut microbiota: *Faecalibacterium prausnizii, Family Ruminococcaceae, Family Clostridiaceae, Family Lachnospiraceae, Order Clostridiales, Bifidobacterium sp.* • 63% (5/8) of the maternal species associated with excessive gestational weight gain was predictive of childhood BMI in the infant gut microbiota: *Faecalibacterium prausnizii, Family Ruminococcaceae, Blautia sp. Bacteroides sp., Bifidobacterium bifidum.*	• The main outcome of the study was to examine the gut microbiota on BMI outcome of the infant.
Pre-pregnancy BMI (or maternal weight status or	(Stanislawski et al., 2017) Norway	Microbiota till 2 y, BMI at 12 y	Cohort	*Confounding factors:* mode of delivery, gestational age at birth, exclusive breastfeeding, antibiotic exposure.	• It is good to note that the study used confounding

Factor	**(Reference)** **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Gut microbiota method**	**Main findings** **Confounding factor** ***1. Alpha diversity, AD*** ***2. Beta diversity/Clusters, BDC*** ***3. Taxa; colour code:*** **Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria** ***4. Other***	**Comments**
gestational weight gain)	Norwegian birth cohort	169 (male = not specified) From 552 children	16S rRNA, V4 region, Illumina Hi-Seq	*1. AD* (Observed OTUs, Shannon Index, FD): • No associations between AD and maternal obesity or AD and excessive gestational weight gain. *3. Taxa* (Relative Abundance): • No associations between maternal obesity or excessive gestational weight gain and infant bacterial taxa.	factors in the statistical models.
Mode of delivery (Vaginal or caesarean)	(Stinson et al., 2018) Review USA, Italy, China, Singapore, Luxembourg. Canada, Denmark, Netherlands 15 studies compared	Range from 0 to 12 mo Range from 2 to 1032 (male = N.A.)	Narrative review 16S rRNA, DGGE, qPCR	*Confounding factors:* antibiotics, labour, breastfeeding, maternal obesity, gestational age, neonatal intensive care unit exposure. *1. AD*: • Caesarean delivered infants had lower AD at 4 mo. *3. Taxa* (Relative Abundance): *Genus:* • Vaginal delivered infants had higher abundance of *Bacteroides spp.* at 1 y and *Parabacteroides spp.* at 4 mo. • Caesarean infants had higher abundance and of *Clostridium spp.* at 1 y and lower abundance of *Escherichia–Shigella spp.* • After adjusting for confounders, mode of delivery had no effect at 6 wk.	• Due to the different inclusion criteria, the 15 studies compared in this review do not include any of the 7 studies from the review in 2016 below. • It is good to note that confounding factors were considered in this review. • The main findings were based on individual studies instead of looking at all the studies as most studies were hard to compare because of the different methodologies used to examine the gut microbiota.

Factor	**(Reference)** **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Gut microbiota method**	**Main findings** **Confounding factor** ***1. Alpha diversity, AD*** ***2. Beta diversity/Clusters, BDC*** ***3. Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria ***4. Other***	**Comments**
Mode of delivery (Vaginal or caesarean)	(Rutayisire et al., 2016) Review Singapore, India, Sweden, Greece, Finland 7 studies included	Range from 0 to 12 mo Range from 24 to 165 (male = N.A.)	Systematic review 16S rRNA, FISH, culture plating	*1. AD*: • Caesarean delivered infants had lower AD at the first 3 mo. • The effect on AD was not seen after 3 mo. *3. Taxa* (Relative Abundance): *Phylum:* • Caesarean had lower abundance of Actinobacteria, and Bacteroidetes, and higher abundance and of Firmicutes at the first 3 mo. *Genus:* • Caesarean had lower abundance of *Bifidobacterium*, and *Bacteroides*, and higher abundance and of *Clostridium* and *Lactobacillus* at the first 3 mo.	• 5 out of 7 of the studies had more than double the number of vaginal delivered infants compared to caesarean. • The review did not look at other confounding effects, and this may be a possible explanation for the effect only seen at 3 mo but not at 6 or 12 mo.
Infant feeding (Breast milk or formula milk)	(Ho et al., 2018) Review Bangladesh, Canada, Haiti, South Africa, USA 7 studies included	Range from 0 to 6 mo Range from 21 to 322 (male = N.A.)	Meta-analysis 16S rRNA	*Confounding factors:* mode of delivery, infant sex, infant age. *1. AD* (Shannon Index, FD, Chao1, Observed OTUs): • Non-EBF or shorter duration of EBF infants had higher AD at 6 mo. *3. Taxa* (Relative Abundance): *Phylum:* • Non-EBF had higher increase in abundance of Bacteroidetes and Firmicutes than EBF.	• It is good to note that less than 6 mo was the point when the authors stopped due to the introduction of solids usually starting by 6 mo. • In addition, sensitivity analyses were done taking into account confounders.

Factor	(Reference) Country Study name if applicable	Participant age Sample size used in microbiota analyses (sex)	Study type Gut microbiota method	Main findings Confounding factor 1. *Alpha diversity, AD* 2. *Beta diversity/Clusters, BDC* 3. *Taxa; colour code:* Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. *Other*	Comments
				Order: • Non-EBF had higher increase in abundance of Bacteriodales and Clostridales than EBF. *Family:* • Non-EBF had higher increase in abundance of Bacteroidaceae and Veillonellaceae than EBF. *Genus:* • Non-EBF had higher increase in abundance of *Bacteroides, Eubacterium, Veillonella* and *Megasphaera* than EBF.	
Antibiotics	(Langdon et al., 2016) Review USA 3 studies included (children) Out of 18 included about antibiotics on gut microbiota.	Range from 0 to 2 y Range from 67 to 160 (male = N.A.)	Narrative review 16S rRNA, culture plating	*1. AD*: • Antibiotic (Ampicillin) associated with decreased AD. • Antibiotic (Cefotaxime) associated with decreased bacterial cell count. *3. Taxa* (Relative Abundance, 16S rRNA; cfu/g, culture plating): *Genus:* • Infants with antibiotic exposure had higher abundance of *Enterobacter spp*. • In contrast, children with antibiotic exposure had lower abundance of *Enterobacteria.*	• The review looked at both children and adults. The findings reported in this table focus on the studies in children. • The contrasting results seen for *Enterobacteria* may be due to the different microbiota method used in the studies. Moreover, different age groups of children were looked at.

Factor	**(Reference)** **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Gut microbiota method**	**Main findings** **Confounding factor** ***1. Alpha diversity, AD*** ***2. Beta diversity/Clusters, BDC*** ***3. Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria ***4. Other***	**Comments**
Parity	(Levin et al., 2016) USA WHEALS birth cohort	0 to 7 mo 298 (male = 157) From 1258	Cohort 16S rRNA, V4 region, Illumina Mi-Seq	*1. AD* (Observed OTUs, Pielou's evenness, FD): • Neonates had +ve association for parity and FD. *2. BDC* (wt/unwt UniFrac): • 0.8% of unwt UniFrac var was explained by parity in infants. *3. Taxa* (Relative Abundance): *Genus:* • Parity was -ve associated with *Clostridium* and *Bacteroides* and +ve association with *Lactobacillus*.	• Similar to this study, many other studies looking at parity looked at many other factors at the same time, i.e. parity was not the main outcome of interest.
Parity	(Laursen et al., 2015) Denmark SKOT I cohort	9 to 18 mo 114 randomly selected samples (male = not specified) From 311	Cross-sectional (2 time points) 16S rRNA, V3 region, Ion OneTouch	*Confounding factors:* mode of delivery, gestational age, infant age, age of start of day-care, breastfeeding duration, daily macronutrient intake. *1. AD* (Shannon Index, Observed OTUs): • Infants with older siblings had higher AD at 18 mo but not at 9 mo. *3. Taxa* (Relative Abundance): *Phylum:* • Numbers of older siblings +ve associated with Firmicutes and Bacteroidetes at 18 mo.	• Study did not explain why they chose to use 114 participants instead of the 227 participants after removing use of antibiotics and improperly stored faecal samples. • It is good to note that confounding factors were considered in this paper. However, they were looked at between the 3 groups with >2, with 1, and without older siblings and

Factor	(Reference) Country Study name if applicable	Participant age Sample size used in microbiota analyses (sex)	Study type Gut microbiota method	Main findings Confounding factor 1. *Alpha diversity, AD* 2. *Beta diversity/Clusters, BDC* 3. *Taxa; colour code:* Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. *Other*	Comments
				Genus: • Presence of older siblings +ve associated with *Haemophilus* and *Faecalibacterium* at 9 mo. • Presence of older siblings +ve associated with *Barnesiella*, *Odoribacter*, *Asaccharobacter* and *Gordonibacter* at 18 mo.	not in the statistical models.

Abbreviations: BMI, body mass index; FISH, Fluorescence in situ hybridization of bacterial cells; DGGE, Denaturing gradient gel electrophoresis; qPCR; quantitative polymerase chain reaction, also known as real-time PCR; FD, Faith's phylogenetic; wt, weighted; unwt, unweighted; var, variance; EBF, exclusively breastfed; -ve, negative; +ve, positive; N.A., not applicable; mo, month.

2.10 How are the diet and gut microbiota associated?

Even though there are many studies looking at the gut microbiota in human adults and animals, a search of the literature using the criteria detailed in section 2.1 found only 14 studies between 2006 and April 2019 in human children that mention dietary components in association with the gut microbiota. These are discussed below (**Table 2.8**). Please note – in this section, colour codes are used to differentiate different bacterial taxa as follows: Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria.

2.10.1 Associations between diet and alpha diversity

In general, alpha diversity is reported to increase with introduction of complementary foods. Two out of the 14 studies (Laursen et al., 2016; Smith-Brown et al., 2016) specifically investigated possible associations between diet and alpha diversity (**Table 2.8**). In one study, **protein**, **fibre**, **cheese**, **meat**, and **rye bread** were positively associated while **fat** was negatively associated with the Shannon Index (Laursen et al., 2016). This was a cross-sectional analysis at 9 months as the 7-day food record (similar to an FFQ) was carried out only at 9 months. Notably, the food record was validated for macronutrients except dietary fibre. As mentioned in section 2.4, it is important to validate the FFQ so that we can be confident of the accuracy of the values obtained from the FFQ. Nonetheless, the positive association with dietary fibre is expected as fibre is a source of food for the gut microbiota.

Consistent with Laursen and colleagues (Laursen et al., 2016), in a large study (n = 1632) in adults, a positive association was found between dietary fibre and Shannon Index which did not change even after adjustment for saturated fat intake (Menni et al., 2017). In contrast, in a meta-analysis on 6 studies in adults, dietary fibre interventions did not have any associations with Shannon Index (So et al., 2018). Menni and colleagues, however, did not find any associations between protein and alpha diversity (Menni et al., 2017). Not many studies have been conducted specifically looking at protein, but those that have mention effects on bacterial taxa instead of alpha diversity (Graf et al., 2015). A possible explanation for the similar associations found for rye bread and protein and dietary fibre may be because rye bread is a traditional Danish food and is one of the main types of Danish breads (Laursen et al., 2016) and whole grains in Danish adults (Helnæs et al., 2016). Hence, a large proportion of the protein and dietary fibre intake could be due to the intake of rye bread, would be in line with

the positive associations of rye bread with Shannon Index, which is similar to that of protein and dietary fibre.

Although Laursen and colleagues found a positive association between **cheese** intake and alpha diversity (Laursen et al., 2016), the other study in young children found that **dairy** was negatively associated with Shannon Index and Chao1 (Smith-Brown et al., 2016). A possible explanation for the difference may be the different age groups looked at, 9 months vs 2 – 3 years. Nine months is just a few months after introduction to complementary feeding while at 2 years, the toddler would be exposed to have been exposed to more types of foods and possibly also drinking more cow's milk (dairy).

For both studies that reported an association between dietary components and alpha diversity (Laursen et al., 2016; Smith-Brown et al., 2016), Spearman's correlations were used to determine these associations. Hence, none of these associations were able to take into account other covariates such as infant feeding, mode of delivery, and parity. This is important because, as mentioned in section 2.9, there may be possible interactions between these factors and diet and alpha diversity.

2.10.2 Associations between diet and beta diversity or clustering

Since the introduction of enterotyping (a method for clustering gut microbiota data) in adults (Arumugam et al., 2011), enterotypes have also been used in 3 out of the 14 studies in children (Nakayama et al., 2015; Nakayama et al., 2017; Zhong et al., 2019) (**Table 2.8**). These enterotypes were named as: E1 *Bacteroides*, E2 *Prevotella*, E3 *Bifidobacterium* (Zhong et al., 2019); and P-type Prevotellaceae, or Bifidobacteriaceae/ Bacteroidaceae BB-type (Nakayama et al., 2015; Nakayama et al., 2017). It seems that E2 is similar to the P-type, with the genus *Prevotella* from the family Prevotellaceae, and E1 and E3 similar to the BB-type as genera *Bacteroides* belongs to family Bacteroidaceae, and genera *Bifidobacterium* belongs to the family Bifidobacteriaceae.

In the study by Zhong et al. (2019), **plant-based protein** and dietary **fibre** intakes were negatively associated with E1 (*Bacteroides*) and E2 (*Prevotella*). The authors did not directly mention the possible reason for the associations and instead mentioned that the children high in E1 (*Bacteroides*), had higher abundance of gut bacterial genes related to butyrate biosynthesis while the children in E2 (*Prevotella*)

had potential for succinate production. This seems contrary to other literature which finds positive associations between dietary fibre and butyrate synthesis as fibre is a food source for bacteria whose end products would be SCFAs such as butyrate (Baxter et al., 2019; Koh et al., 2016). *Bacteroides* species are known propionate produces while members of the Firmicutes phylum are known butyrate producers (Rowland et al., 2018). Hence, bacterial cross-feeding (Flint, Scott, Duncan, Louis, & Forano, 2012a) would be a reasonable rationale for the finding that the children high in E1 was associated with butyrate biosynthesis. **Fat** intake was positively associated with BB-type (Bifidobacteriaceae/ Bacteroidaceae) while **β-carotene** and **vitamin A** were positively associated with P-type (Prevotellaceae) (Nakayama et al., 2017). In addition, **rice**, **chicken**, **soy** and **eggs** were found to be positively associated with P-type (Prevotellaceae) (Nakayama et al., 2015). It is important to note that only frequencies and not actual amounts of the food items were collected (Nakayama et al., 2015). Moreover, the study looked mostly at microbiota differences between countries rather than associations between diet and enterotype. The study relates resistant starch to rice intake as an explanation for the association with P-type (Prevotellaceae). However, resistant starch values were not calculated in the study. In addition, the resistant starch value of rice is 0.37 g/100 g which is not high compared to other possible sources of resistant starch such as potatoes (baked, boiled, mashed, instant (hot)) at 0.59 g/100 g (Landon et al., 2012) which was in one of the food groups that the study investigated. Moreover, resistant starch has different forms, RS1 (physically inaccessible starch), RS2 (resistant starch granules), RS3 (retrograded starch) and RS4 (chemically modified starch), with *Ruminococcus bromii* (of family Ruminococcaceae) being a keystone resistant starch degrader (Vital et al., 2018).

Another popular method of clustering and summarising data is the PCoA plot. Following PCoA, statistics such as PERMANOVA are then used to determine whether there are differences between the clusters. Different beta diversity measures such as weighted or unweighted UniFrac or Bray Curtis measures can be used in the PCoA plot which will result in different interpretations of the results. This is because Bray Curtis looks at how different the samples are by distance matrix while UniFrac considers the phylogenetic distance between the sets of bacteria as formed in a phylogenetic tree. In one RCT, no clear differences between the different dietary interventions (**cereal** vs

cereal + fruit vs **meat**) were found using a PCoA plot with unweighted UniFrac and PERMANOVA statistics (Qasem et al., 2017) (Section 2.8.2c).

Dietary associations in the other studies were generally inferred as the PCoA clustered the participants by geographical locations, and diet was associated with the different groups by locations (De Filippo et al., 2017; Ruggles et al., 2018).

Another study used Adonis (a non-parametric statistical method that identifies relevant clusters of the data and calculates the squared deviations from the centres of these clusters) with a weighted UniFrac measure and found that **yoghurt** explained 9%, **milk alternatives** explained 6% and **soy products** explained 7% of the variance of the association between the different dietary components and beta diversity (weighted UniFrac) (Smith-Brown et al., 2016).

2.10.3 Associations between diet and bacterial taxa

It is difficult to tease out the different dietary components and their related associations with the different bacterial taxa of the gut microbiota as studies report different dietary components and there are also thousands of bacterial taxa that can be looked at, from the phylum all the way to the species level.

Fat (Nakayama et al., 2017) and **dairy** (Smith-Brown et al., 2016) have been positively associated while **fruit** (Smith-Brown et al., 2016) has been negatively associated with F/B ratio in studies in children. The F/B ratio has generally been associated with obesity, with a higher F/B ratio associated with obese participants (although the results are inconsistent and inconclusive, as reported in a systematic review looking at studies in adults (Castaner et al., 2018)).

Intake of **dietary fibre** has been reported to have positive associations with genera *Prevotella* (De Filippo et al., 2017) and Xylanibacter (De Filippo et al., 2010), and families Eubacteriaceae, Pasteurellaceae, Prevotellaceae and Veillonellaceae (Laursen et al., 2016); and negative associations with families Bifidobacteriaceae, Enterococcaceae and Lactobacillaceae (Laursen et al., 2016). Looking at carbohydrates, **rice** was negatively associated with *Lactobacillus*, *Bacteroides fragilis* and *Prevotella* while **bread** was negatively associated with *Bifidobacterium* (La-Ongkham et al., 2015). **Fruits** have a negative association with the family Erysipelotrichaceae (Smith-Brown et al., 2016) and a positive association *Lactobacillus* (La-Ongkham et al., 2015).

Similarly, **vegetables** are positively associated with genera *Lachnospira* (Smith-Brown et al., 2016), *Lactobacillus*, *Eubacterium rectale* and *Prevotella* (La-Ongkham et al., 2015) and negatively associated with *Clostridiales* (Smith-Brown et al., 2016). Fruits, vegetables and carbohydrate-based foods are good sources of dietary fibre and fibre fractions which are major food sources for the gut bacteria. In-vitro and animal studies have shown that different fibre fractions may promote more of certain bacterial taxa, for example pectin which can be found in fruits promotes Bifidobacteria and Lactobacilli while inulin (found in fruits and vegetables) increases *Catenibacterium* and *Blautia* (Williams, Grant, Gidley, & Mikkelsen, 2017). In addition, in non-westernised populations, intake of dietary fibre drives abundance of *Prevotella* (Makki et al., 2018). Resistant starch which can be found in many carbohydrate-based foods, enriches *Bifidobacterium adolescentis*, *R. bromii*, *Eubacterium rectale*, and *Parabacteroides distasonis* in adults (Martínez, Kim, Duffy, Schlegel, & Walter, 2010). Moreover, *R. bromii* is recognised as a keystone species for resistant starch degradation as compared to the other amylolytic bacteria such as *E. rectale* and *B. adolescentis* (Ze, Duncan, Louis, & Flint, 2012). This in part can be contributed to the unique organization of *R. bromii*, which has extracellular starch-degrading enzymes attached to the cell surface, also referred to as 'amylosomes' (Ze et al., 2015).

Vegetarian protein was found to be negatively associated with the phylum Firmicutes (Smith-Brown et al., 2016). **Protein** was negatively associated with the families Bifidobacteriaceae, Enterococcaceae, Lactobacillaceae while positively associated with the families Erysipelotrichaceae, Peptostreptococcaceae, Lachnospiraceae, Clostridiaceae, Sutterellaceae, Ruminococcaceae (Laursen et al., 2016) and genera *Bacteroides*, *Parabacteroides* and *Oscillibacter* (Nakayama et al., 2017). Considering foods that are high in protein, **chicken** was positively associated with *Clostridium leptum*, **beef** negatively associated with *Bifidobacterium*, and **fish** negatively associated with *Clostridium leptum* and *Bifidobacterium* (La-Ongkham et al., 2015). In addition, meat in an RCT was found to be associated with an overrepresentation of the family Enterobacteriaceae (Qasem et al., 2017). Protein is linked to the gut microbiota by the fermentation of amino-acids (from the digestion of proteins by enzymes produced by the human host), which would bring about metabolically active compounds such as SCFAs, branch chained fatty acids or different nitrogen containing compounds (Lin, Liu, Piao, & Zhu, 2017). In particular, the

Clostridium genus for lysine or proline, and *Peptostreptococcus* genus for glutamate or tryptophan (Lin et al., 2017). In addition, in a narrative review that compared plant and animal protein, plant protein was associated with increased *Bifidobacterium*, and *Lactobacillus* and decreased *Bacteroides* and *Clostridium perfringens* which is associated with increased SCFAs, and hence a positive effect with increased gut barrier (Singh et al., 2017). In contrast, animal protein was associated with increased *Bacteroides*, *Alistipes*, *Bilophila* and *Ruminococcus* and decreased *Bifidobacterium* which is associated with decreased SCFAs, and a negative effect on cardiovascular disease (Singh et al., 2017). *Clostridium*, *Peptostreptococci*, and *Bacteroides* have been found to be proteolytic fermenters in *in vitro* studies (Diether & Willing, 2019). Deleterious nitrogenous compounds such as genotoxic nitrosamines can be produced from protein fermentation (Gratz, Wallace, & El-Nezami, 2011), and long-term intake of high protein diets may be detrimental to colonic health (Diether & Willing, 2019; Russell et al., 2011).

Dairy has been positively associated with *Bifidobacterium* (De Filippo et al., 2017), *Streptococcus*, *Lachnoclostridium* and *Erysipelatoclostridium* while negatively associated with *Fusicatenibacter*, *Faecalibacterium prausnitzii* , *Parabacteroides* and the phylum Bacteroidetes (Smith-Brown et al., 2016). This is in alignment with studies which show that *F. prausnitzii* growth is promoted by prebiotics such as fructooligosaccharides (FOS) (Scott, Martin, Duncan, & Flint, 2014). **Yoghurt** has been positively associated with *Streptococcus*, while negatively associated with *Alistipes* and *Bacteroides* (Smith-Brown et al., 2016). Yoghurt is a fermented dairy product which is usually fermented with lactic acid producing bacteria, however, the study did not specifically mention whether the yoghurt intake was 'live' yoghurt. *Streptococcus* species is part of lactic acid producing bacteria and have been found to be important in playing a role in preventing pathogens from colonizing the gut (Adolfsson, Meydani, & Russell, 2004).

Fat was negatively associated with *Prevotella* while positively associated with *Bacteroides* and the order *Clostridiales* (Nakayama et al., 2017). A Westernised diet, defined as a high fat, high protein and low dietary fibre diet (Marchesi et al., 2016), has been associated with lower *Prevotella* and higher *Bacteroides* and *Clostridiales species* (Gorvitovskaia, Holmes, & Huse, 2016).

Two RCTs have investigated the impact of **iron** supplementation on children's microbiota. In the first study, supplementation resulted in overall more *Clostridium* and Escherichia and less *Bifidobacterium* (Jaeggi et al., 2015). This is supported in part by a study that looked at iron fortification of infant formula, and found that consumption of a high-iron infant formula was associated with a lower abundance of Bifidobacteria compared to the low-iron infant formula (Simonyté Sjödin et al., 2018). The second study also found an association with Enterobacteriaceae with dietary iron (the direction of association was not mentioned) (Krebs et al., 2013). Decreasing *Bifidobacterium* which is a known beneficial bacterium may have negative effects on the gut microbiota community due to decreasing 'colonization resistance' (Donaldson et al., 2016). This is supported in part by an in-vitro study which showed that increased iron availability is associated with increased pathogenic bacteria such as *Escherichia coli* (Parmanand et al., 2019).

Table 2.8: Studies (n = 14) examining effects of diet on the gut microbiota from start of complementary feeding till 11 years of age.

(Reference) **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Dietary method** **Gut microbiota method**	**Main findings (Diet with faecal gut microbiota)** ***1. Alpha diversity, AD*** ***2. Beta diversity/Clusters, BDC*** ***3. Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria ***4. Other***	**Stat used and confounders adj for**	**Comments**
(Zhong et al., 2019) Netherlands KOALA Birth Cohort Study	6 – 9 y 281 (male = 139) Sample from larger cohort of 2834	Cross-sectional FFQ at 4-5 y: 71 items over the past 4 wks, using FFQ from Dutman et al. (2011) qPCR, region not specified, Illumina, and shotgun metagenomic sequencing	*1. AD* (Shannon Index): • E3 had lower alpha diversity. *2. BDC* (Enterotypes): • 281 participants were separated into 3 enterotypes: *Bacteroides* E1 (n = 143). *Prevotella* E2 (n = 74). *Bifidobacterium* E3 (n = 64). • Plant-based protein and fibre -ve with E1 and E2 **[*p*<0.05].**	*AD:* Wilcoxon rank-sum test and Kruskal-Wallis test, adj= BH corr *Enterotypes*: DMM and PAM-clustering using JSD and Bray-Curtis, CH index at the genus level. PERMANOVA, R 'adonis' on 9999 permutations [*p-value*] GLM regression, confounders of gender, age, BMI-z score and early events.	• The FFQ was designed and validated only for energy intake. • It is good to note that recent antibiotic use was an exclusion criterion for this study, and that confounders were considered. • Not much focus on the diet and microbiota taxa, with more focus on diet and the diff enterotypes.

(Reference) **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Dietary method** **Gut microbiota method**	**Main findings (Diet with faecal gut microbiota)** ***1. Alpha diversity, AD*** ***2. Beta diversity/Clusters, BDC*** ***3. Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria ***4. Other***	**Stat used and confounders adj for**	**Comments**
			4. Other: E3 showed higher enrichment of functions involved in metabolism of simple sugars including glycolysis and the pentose-phosphate pathway.	Residuals checked using Shapiro-Wilk *Other*: KEGG, Wilcoxon rank-sum test, adj= BH corr	
(Ruggles et al., 2018) USA and Venezuela	2 – 8 y 24 (male = not specified) 2: USA (visitors) 22: Venezuela (villagers) Sample from a mixture with adults of 45	Comparative (16-d experiment) Not measured (Inferred to be an Amerindian diet; traditional low fat/high-fibre unprocessed diet) 16S rRNA, V4 region, Illumina Mi-Seq	*1. AD* (FD & Shannon Index & Observed OTUs): Results presented as boxplot and only FD mentioned. • Visitors had lower FD than Villagers **[*p*=0.020]**. • Visitors increased in FD after 16-d **[*p*=not specified?]**. *2. BDC* (unwt/wt UniFrac & Bray-Curtis): • Visitors had diff unwt UniFrac than Villagers **[*p*<0.001]**. • Visitors unwt UniFrac changed over time **[*p*=not specified?, observation made from PCoA plot?]**. *3. Taxa* (Relative Abundance): *Genus*: • Visitors had lower *Escherichia* and of unknown Ruminococcaceae and *Bacteroidales* and higher	*AD*: Kruskal-Wallis test with the *post hoc* Dunn's test, adj= BH corr [*p-value*] *BDC*: PCoA (unwt UniFrac), stat= ANOSIM, R pkg, 'vegan', with 999 permutations [*p-value*] *Taxa* (detect overrepresent bacterial taxa):	• Study also looked at microbiota from other sites such as skin, oral and nasal. • Small sample size for multiple comparisons. Hence depending on the number of comparisons, the results (for example a genus) may be significant due to chance and not a real diff. • Visitor samples collected 5-8 times while Villagers' samples only once. • Not mentioned whether the use of antibiotics was looked at or other factors just as sex was matched

(Reference) Country Study name if applicable	Participant age Sample size used in microbiota analyses (sex)	Study type Dietary method Gut microbiota method	Main findings (Diet with faecal gut microbiota) 1. ***Alpha diversity, AD*** 2. ***Beta diversity/Clusters, BDC*** 3. ***Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. ***Other***	Stat used and confounders adj for	Comments
			Bacteroides, *Faecalibacterium*, *Blautia*, *Clostridium*, *Coprococcus*, *Ruminococcus*, *Lachnospira*, *Bifidobacterium*, and unknown Lachnospiraceae and Rikenellaceae than the Villagers **[*p<0.05, not specified*]**. • Visitors increased *Faecalibacterium* after 16-d **[*p<0.05, not specified*]**. *4. Other*: • Visitors showed same 'microbiota age' compared to the Villagers.	Heatmap, LEfSe, LDA score >2.0, stat= Kruskal-Wallis test with the *post hoc* Dunn's test, adj= BH corr [*p-value*] *Other*: Random forest model	between the 2 study samples. • Diet not measured, just observed.
(De Filippo et al., 2017) Africa Italy	2 – 8 y 37 (male = 23) 11: Boulpon, Africa, BR (Rural) 8: Nanoro, Africa BT (Rural) 5: City of Burkina Faso, Africa, BC (City) 13: Florence, Italy, EU (City)	Cross-sectional 4-d dietary questionnaire (Italy) and in-depth interview on diet (Africa) 16S rRNA, V5-V6 region, Roche 454	*1. AD* (Chao1 & Shannon Index & Observed OTUs): • Observed OTUs & Chao1: BT>BR>BC>EU • Shannon Index least in BR compared to the rest *2. BDC* (unwt/wt UniFrac & Bray-Curtis): • PCoA analysis showed a clear diff between BR and EU **[*p* = 0.0001]**. *3. Taxa* (Relative Abundance): *F/B*: clear separation between BR and EU	*AD*: R 'phyoseq', Kruskal–Wallis test or pairwise Wilcoxon test, PERMANOVA *BDC*: PCoA, stat= PERMANOVA *BDC* (taxa): stat= Wilcoxon rank-sum test	• It is good to note that past 6 mos antibiotic and probiotic use and hospitalization were exclusion criteria for this study. • Africa has both city and rural participants, but Italy only has city participants. Hence, comparisons between countries may not be comparable. • Study mentioned that participants were breastfed

(Reference) Country Study name if applicable	Participant age Sample size used in microbiota analyses (sex)	Study type Dietary method Gut microbiota method	Main findings (Diet with faecal gut microbiota) 1. *Alpha diversity, AD* 2. *Beta diversity/Clusters, BDC* 3. *Taxa; colour code:* Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. *Other*	Stat used and confounders adj for	Comments
	BR and EU from De Filippo et al. (2010)		*Phylum*: • Bacteroidetes BR **[69%]** >BT **[48%]** >BC **[33%]** >EU **[26%]** • Firmicutes BR<BT<BC **[56%]** <EU **[60%]** *Genus*: • *Prevotella* BR **[39%]** >BT>BC **[10%]** >EU: reflecting the reduction of fibre intake. • *Uncultured Lachnospiraceae, Roseburia, Dorea* less in BR and BT and more in BC and EU: Urbanization and westernized diet led to the loss of traditional and rural populations in BC. • *Bacteroides* more in BC and EU children compared with BR and BT: associated with lipids, cholesterol and amino acids intake, and dairy consumption. • *Bifidobacterium*, more in BC and EU children compared with BR and BT: associated with milk and dairy consumption. *4. Other*: • BC enrichment of KEGG categories involved in starch and sucrose metabolism and methane metabolism, related to the fermentation of polysaccharides. • EU low levels of SCFAs, especially butyric acid, compared to BC, BR, BT	[%Relative Abundance] *Other*: PICRUSt by KEGG and SCFAs	(except 1 EU) but did not mention whether they were EBF and the duration of breastfeeding. • Different methods of dietary assessment within the same study may not be comparable. • Study mainly compares the diff in gut microbiota composition between the 4 locations. Diet was not used in the analyses to determine associations with the microbiota but rather inferred from comparing between the 4 locations.

(Reference) **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Dietary method** **Gut microbiota method**	**Main findings (Diet with faecal gut microbiota)** ***1. Alpha diversity, AD*** ***2. Beta diversity/Clusters, BDC*** ***3. Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria ***4. Other***	**Stat used and confounders adj for**	**Comments**
(Qasem et al., 2017) Canada	4 – 6 mos 56 (male = not specified) 18: Cer 19: Cer + F 19: M Sample from initial enrolment of 90	RCT (Iron-fortified cereal vs Iron-fortified cereal + Fruit vs Meat; Cer vs Cer + F vs M) 3-d food records 16S rRNA, V3-V4 region, Illumina Mi-Seq	*1. AD* (Chao1 & Shannon Index): • After introduction of Cer + F **[43.1 (2.1, 84.2)]** or M **[72.9 (32.2, 113.7)]**, +ve Chao1 change. *2. BDC* (unwt UniFrac): • No clear diff in microbiota community structures between Cer vs Cer + F vs M **[*p*=0.22]**. *3. Taxa* (Relative Abundance): Note none of the comparisons were statistically significant for before and after Cer, Cer + F and M. After adj = FDR. *Family*: • After Cer, decreased Bifidobacteriacea **[From 51% to 37%]**. • After meat, overrepresentation of Enterobacteriaceae **[10% in M vs 5% in Cer]**. *4. Other*: • Increase of faecal ROS formation after the introduction of complementary feeding **[*p*<0.002]**. • Increase of faecal calprotectin after the introduction of complementary feeding **[*p*=0.004]**.	*AD*: ANOVA [mean (95% CI)] or 2-sided t-test *BDC*: PCoA, stat= PERMANOVA with 500 permutations [*p-value*] *Taxa* (on dominant taxa >0% median relative abundance): Wilcoxon rank sum test, adj=FDR [% relative abundance] *Other*: rep measures ANOVA [*p-value*]	• Small sample size for multiple comparisons. • From this study, it seems that diet (Cer vs Cer +F vs M) does not play a huge role in gut microbiota changes. Rather, diff before and after introduction of complementary foods are seen. • It is good to note only full term healthy EBF infants were included for this study as infant milk feeding is a well-established factor in affecting the infant gut microbiota.

(Reference) **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Dietary method** **Gut microbiota method**	**Main findings (Diet with faecal gut microbiota)** 1. ***Alpha diversity, AD*** 2. ***Beta diversity/Clusters, BDC*** 3. ***Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. ***Other***	**Stat used and confounders adj for**	**Comments**
(Nakayama et al., 2017) Philippines Asian Microbiome Project	7 – 9 y 43 (male = 26) 24: Baybay (Rural) 19: Ormoc (City) & 295 out of the 303 from a previous study by Nakayama et al. (2015)	Cross-sectional FFQ: 136 items over the past 3 mo, modified from Singapore National Dietary Survey (Singapore Health Promotion Board, 2010) 16S rRNA, V6-V8 region, Roche 454	*1. AD* (FD & Shannon Index & Observed OTUs): • Ormoc had lower FD than Baybay **[*p*=0.0314]**. • BB-type had lower FD **[*p*=0.0193]** and observed OTUs **[*p*=0.0073]** than P-type. *2. BDC* (enterotype): • 43 participants were separated into 2 enterotypes: Prevotellaceae P-type (n = 25). Baybay 87.5%. Bacteroidaceae BB-type (n = 18). Ormoc 78.9%. But also high in Ruminococcaceae and Lachnospiraceae. *2. BDC* (Bray-curtis): • Diff in gut microbiota **[13.1%]** could be ascribed to dietary diff associated with the city of residence **[9.8%]** (CAP plot). • Fat favours BB-type characterized by high abundance of Firmicutes and were opposed to P-type characterized by high abundance of Bacteroidetes (CAP plot). • β-carotene and vitamin A favours P-type (CAP plot).	*AD*: Wilcoxon rank sum test (box-plot) [*p-value*] *Enterotypes*: PCA using 338 participants at family level taxa in R 'EMBL'; 'prcomp', CH index, silhouette index, 's.class', 'ade4' *Bray-curtis*: CAP using db-RDA on (n = 40) with 999 permutations. [var explained] R 'envfit'. (plots) [*p-value*]	• 2 participants had antibiotics 2 wks prior and were excluded from some analyses. • Original FFQ had 397 items. Fewer than half of the items were used in the study and there wasn't any mention of which items were used. • FFQ not validated in current population. Common food items eaten in Philippines will be different from Singapore. • Good to see the dietary nutrient intake looked at. 3 participants were excluded for dietary analyses (n = 40) but included for other analyses. • Small sample size for multiple comparisons

(Reference) **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Dietary method** **Gut microbiota method**	**Main findings (Diet with faecal gut microbiota)** ***1. Alpha diversity, AD*** ***2. Beta diversity/Clusters, BDC*** ***3. Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria ***4. Other***	**Stat used and confounders adj for**	**Comments**
			3. Taxa (Relative Abundance): *F/B*: Fat +ve **[rho=0.58]**. *Genus*: • Fat -ve with *Prevotella* **[rho = −0.6315; CI95%: −0.79 to −0.40]**, +ve with *Bacteroides* **[rho = 0.5298; CI95%: 0.26 to 0.72]**, +ve with the order *Clostridiales* **[rho = 0.5443; CI95%: 0.28 to 0.73]**. • Dietary fat and protein commonly associated with a number of species belonging to genera *Bacteroides*, *Parabacteroides and Oscillibacter*. *4. Other*: • BB-type is metabolically more active with simple sugars, amino acids, and lipids while P-type is more involved in digestion of complex CHO.	*Taxa*: Spearman's [rho] *Other*: PICRUSt by KEGG	
(Laursen et al., 2016) Denmark SKOT I and SKOT II	9 – 18 mos 227 (male = 114) 114: SKOT I 113: SKOT II Sample from larger cohort of 311 SKOT I & 208 SKOT II	Cohort 7-d food record at 9 mo, validated from Gondolf et al. (2012) 16S rRNA, V3 region, Ion OneTouch	*1. AD* (Shannon Index, observed OTUs, Pielou's evenness index): • Protein and fibre +ve with Shannon • Fat -ve with Shannon • Family food dietary pattern +ve with Shannon • Cheese **[rho=0.30]**, meat **[rho=0.27]**, and rye bread **[rho=0.24]** +ve with Shannon *2. BDC* (Bray Curtis):	*AD*: Spearman's [rho], adj=FDR *BDC*: PCoA	• 7-d food record was validated for macronutrients but not fibre. • Good that current antibiotic use was an exclusion criterion, and that past antibiotic used (2 wk before faecal sample) was investigated.

(Reference) Country Study name if applicable	Participant age Sample size used in microbiota analyses (sex)	Study type Dietary method Gut microbiota method	Main findings (Diet with faecal gut microbiota) 1. *Alpha diversity, AD* 2. *Beta diversity/Clusters, BDC* 3. *Taxa; colour code:* Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. *Other*	Stat used and confounders adj for	Comments
			• Beta diversity clustered according to age rather than cohort. *3. Taxa* (Relative Abundance): *Family*: • Bifidobacteriaceae, Enterococcaceae, Lactobacillaceae -ve with fibre and protein. • Erysipelotrichaceae, Peptostreptococcaceae, Lachnospiraceae, Clostridiaceae, Sutterellaceae, Ruminococcaceae +ve with protein. • Eubacteriaceae, Pasteurellaceae, Prevotellaceae, Veillonellaceae +ve with fibre.	*Taxa*: Heatmap Spearman's, adj=FDR	• It would be better to use other statistical methods that could take into account other confounding factors that may affect the gut microbiota.
(Smith-Brown et al., 2016) Australia Feeding Queensland Babies	2 – 3 y 37 (male = 21) Sample from larger cohort of 462	Cross-sectional • FFQ: 120 items over the past 6 mo, validated from Watson, Collins, Sibbritt, Dibley, and Garg (2009) • 24-hour recall: 3 days before 16S rRNA, region not specified, Illumina Mi-Seq	*1. AD* (Chao1 & Shannon Index): • Dairy (FFQ) -ve with Shannon **[rho=-0.45]**, Dairy (24hr recall) -ve with Chao1 **[rho=-0.51]**. • Fruit **[rho=0.36]** and vegetarian protein **[rho=0.36]** (FFQ) +ve with Shannon, but not sig after correction for FDR. *2. BDC* (wt UniFrac): • Yoghurt (FFQ) explained 9% of the var **[R^2=0.09]**. • Milk alternatives (FFQ) 6% of the var **[R^2=0.06]**. • Soy products (FFQ) 7% of the var **[R^2=0.07]**. *3. Taxa* (Relative Abundance): *F/B*: • Dairy (FFQ) +ve **[rho=0.47]**. • Fruit (FFQ) -ve **[rho=-0.40]**.	*AD*: Spearman's [rho], adj=FDR *BDC*: Adonis [R^2], adj=FDR *Taxa* (on those associated with *BDC*):	• Small sample size for multiple comparisons. • The FFQ was developed and validated for youth aged 9-16 years. • Not specified how many 24-hour recalls were done per participant. • Not specified which region was used for the 16S rRNA sequencing. • The directions of the associations for both FFQ and Shannon Index were in the same direction although different dietary

(Reference) Country Study name if applicable	Participant age Sample size used in microbiota analyses (sex)	Study type Dietary method Gut microbiota method	Main findings (Diet with faecal gut microbiota) 1. *Alpha diversity, AD* 2. *Beta diversity/Clusters, BDC* 3. *Taxa; colour code:* Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. *Other*	Stat used and confounders adj for	Comments
			Phylum: • Dairy (FFQ) -ve with Bacteroidetes **[rho=-0.48]**. • Vegetarian protein (FFQ) -ve with Firmicutes **[rho=-0.44]**. *Genus*: • Yoghurt (FFQ) **[rho=0.65]** and Dairy (24hr recall) **[rho=0.58]** +ve with *Streptococcus*. • Dairy (24hr recall) +ve with *Lachnoclostridium* **[rho=0.51]** and *Erysipelatoclostridium* **[rho=0.51]**, and -ve with *Fusicatenibacter* **[rho=-0.51]**. • Yoghurt (FFQ) -ve with *Alistipes* **[rho=-0.62]** and *Bacteroides* **[rho=-0.51]**. • Dairy (24hr recall) -ve with *Parabacteroides* **[rho=-0.51]**. • Vegetable (24hr recall) +ve with *Lachnospira* **[rho=0.49]** and -ve with *Clostridiales* **[rho=-0.51]**. • Fruit (FFQ) -ve with unspecified genus of family Erysipelotrichaceae **[rho=-0.51]**. *Species*: • Dairy (24hr recall) -ve with *Faecalibacterium prausnitzii* **[rho=-0.58]**. *4. Other*: Fruit (FFQ) +ve with digestive system level 2 KEGG functional pathway **[rho=0.58]**.	Spearman's [rho], adj=FDR *Other*: Spearman's [rho], adj=FDR	components were significantly associated with microbiota outcomes for different dietary methods. • It is good to note that recent antibiotic and medications use were exclusion criteria for this study. It would be better to use other statistical methods that could take into account other confounding factors that may affect the gut microbiota.

(Reference) **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Dietary method** **Gut microbiota method**	**Main findings (Diet with faecal gut microbiota)** 1. ***Alpha diversity, AD*** 2. ***Beta diversity/Clusters, BDC*** 3. ***Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. ***Other***	**Stat used and confounders adj for**	**Comments**
(Nakayama et al., 2015) China Japan Taiwan Thailand Indonesia Asian Microbiome Project	7 – 11 y 303 (male = 159) Rural & City 59: China 83: Japan 53: Taiwan (no FFQ) 52: Thailand 56: Indonesia	Cross-sectional FFQ: qualitative, 11 items over the past 2 wks. (No other details specified.) 16S rRNA, V6-V8 region, Roche 454	*1. AD* (FD & Shannon Index & Observed OTUs): • Indonesia showed the highest AD, which is common property of the P-type. • China showed high AD although associated with the BB-type. • Japan showed low AD. *2. BDC* (enterotype): • 303 participants were separated into 2 enterotypes: Prevotellaceae P-type. Bifidobacteriaceae/ Bacteroidaceae BB-type. Also high in Ruminococcaceae and Lachnospiraceae. *Enterotypes & diet*: • Rice +ve with P-type **[1.79 (1.29, 2.48)]**. • Chicken **[1.96 (1.36, 2.83)]**, soy **[2.08 (1.45, 2.98)]** and eggs **[2.25 (1.61, 2.62)]** +ve with P-type. *3. Taxa* (Relative Abundance): *Confounding (Genus):* *Bifidobacterium* **[-1.75%/y]** and two Lachnospiraceae genera, *Blautia* **[11.22%/y]** and *Roseburia* **[10.72%/y]**, associated with age. But no change in the age-adj multivariate analysis.	*AD*: Wilcoxon rank sum test (box-plot) *Enterotypes*: PCA using 303 participants at family level taxa in R 'EMBL'; 'prcomp', CH index, silhouette index, 's.class', 'ade4' *Enterotypes & diet*: Log regression [odds ratio, (95%CI)] *Confounding:* Age bias between cities-multivariate analysis [%change/y]	• Dietary assessment method not described in detail. • Study looked mostly at microbiota diff between countries instead of diet effects. • Effects of diet on bacterial taxa was not mentioned, only effects on the enterotypes. • Good that confounding effects of age was looked at. • However, no mention about recent antibiotic and medications usage.

(Reference) **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Dietary method** **Gut microbiota method**	**Main findings (Diet with faecal gut microbiota)** 1. ***Alpha diversity, AD*** 2. ***Beta diversity/Clusters, BDC*** 3. ***Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. ***Other***	**Stat used and confounders adj for**	**Comments**
			4. Other: P-type has enrichment of amylases and decrease of glucosidase.	*Other*: PICRUSt by KEGG	
(La-Ongkham et al., 2015) Thailand	8 – 11 y 60 (male = 28) 29: Khon Kaen (Rural) 31: Bangkok (City)	Cross-sectional FFQ: qualitative, 14 items over the past 1 mo. (Not validated) qPCR, 16S rRNA, V6-V8 region	*3. Taxa* (log copy numbers per wet gram faeces): • Chicken +ve with *Clostridium leptum* **[0.29]**. • Beef **[-0.33]** and fish **[-0.41]** -ve with *Bifidobacterium.* • Fish -ve with *Clostridium leptum* **[-0.26]**. • Rice -ve with *Lactobacillus* **[-0.26]** and *Bacteroides fragilis* **[-0.27]** and *Prevotella* **[-0.26]**. • Bread -ve with *Bifidobacterium* **[-0.29]**. • Fruit +ve with *Lactobacillus* **[0.29].** • Vegetable +ve with *Lactobacillus* **[0.42]**, *Eubacterium rectale* **[0.26]** and *Prevotella* **[0.34]**.	*Taxa:* Spearman's [rho]	• Exclusion criteria included: gastrointestinal disease and any diarrhoea or taking antibiotics in the past 1 mo. • Only frequencies for diet and very few food items captured.
(Jaeggi et al., 2015) Africa	5 – 6 mos 101 (male = not specified) 28: +2.5mgFeMNP 21: -2.5mgFeMNP 26:+12.5mgFeMNP 26:-12.5mgFeMNP	RCT (3 wk) FeMNP: Iron micronutrient powders Biochemical iron indices qPCR, 16s rRNA V3-V6 region, Roche 454	*3. Taxa* (Relative Abundance): *Genus:* • +FeMNP more *Clostridium* **[*p*=0.033]** and Escherichia **[*p*=0.010]** and less *Bifidobacterium* **[*p*=0.085]**.	*Taxa:* Mann-Whitney U, GLM [*p-value*] *Confounding*: Gender, season in which the intervention started, start date of CF (Baseline values)	• It is good that confounders are looked at and placed as covariates in the GLM. • Study just looked at iron supplementation on gut microbiota. • Good to note that compliance to RCT intervention was assessed. However, other dietary factors were not considered.

(Reference) **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Dietary method** **Gut microbiota method**	**Main findings (Diet with faecal gut microbiota)** ***1. Alpha diversity, AD*** ***2. Beta diversity/Clusters, BDC*** ***3. Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria ***4. Other***	**Stat used and confounders adj for**	**Comments**
			4. Other: No sig diff in faecal acetate, propionate, or butyrate concentrations between +FeMNPs and −FeMNPs.	*Other:* SCFAs	
(Bernal et al., 2013) Spain	6 – 10 mos 19 (male = not specified) 10: Cereal A 9: Cereal B	RCT (2 mos) Cereal A higher ratio of complex/simple CHO than Cereal B Daily food intake for 2 mos. qPCR, microbiological agar plates (culture dependent method)	*3. Taxa* (log CFU per g fresh faeces): *Genus:* • Cereal A higher *Bifidobacterium* than Cereal B **[*p*=0.085]**. *4. Other:* Cereal A higher butyrate concentrations than Cereal B.	*Taxa:* Mann-Whitney U [*p-value*] *Other:* SCFAs stat= ANOVA	• Study inclusion criteria included exclusive infant formula feeding since at least 4 mos. • Cereal A had higher digestible starch and resistant starch, lower digestible CHO and both have similar fibre. • No mention about diet results or compliance to RCT cereal intake or antibiotic use.
(Krebs et al., 2013) USA	6 – 10 mos 14 (male = not specified) 6: IZFC 4: IFC 4: M Sample from larger study of 45	RCT (4 mos) Iron and zinc fortified cereal (IZFC) vs iron fortified cereal (IFC) vs meat (M) 3d diet records every mo 16S rRNA, V1-V3 region, Roche 454	*1. AD* (Chao1 & Shannon Index & Observed OTUs): • Increased in OTUs with age (diff between groups were not mentioned) *3. Taxa* (Relative Abundance): *Family:* • Enterobacteriaceae **[*p*=0.03]** sig correlated with dietary iron (direction of association not indicated).	*AD:* Interquartile range *Taxa:* multivariate ANOVA [*p-value*]	• Iron intake only estimated from CF, breast milk not taken into account. Diet results not presented. • Small sample size may not be powered for multiple comparisons.

(Reference) **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Dietary method** **Gut microbiota method**	**Main findings (Diet with faecal gut microbiota)** 1. ***Alpha diversity, AD*** 2. ***Beta diversity/Clusters, BDC*** 3. ***Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. ***Other***	**Stat used and confounders adj for**	**Comments**
			Genus: • *Bifidobacterium* **[*p*=0.004]** and *Rothia* **[*p*=0.007]** decreased in IFC over time but remained unchanged for IZFC and M. • *order Lactobacillales* **[*p*=0.04]** decreased in IFC over time compared to IZFC and M. • *order Bacteroidales* **[*p*=0.02]** more abundant in IFC over time compared to IZFC and M.		
(De Filippo et al., 2010) Africa Italy	1 – 6 y 29 (male = 18) 14: BF Burkina Faso (Rural) 15: EU Florence, Italy (City)	Cross-sectional 3-d dietary questionnaire (Italy) and in-depth interview on diet (Africa) 16S rRNA, V5-V6 region, Roche 454	*1. AD* (Chao1 & Shannon Index): • BF higher Chao1 and Shannon than EU. *3. Taxa* (Relative Abundance): *F/B:* EU **[2.8±0.06]** greater than BF **[0.47±0.05].** *Phylum:* • Actinobacteria more in BF **[10%]** than EU **[7%]**. • Bacteroidetes more in BF **[58%]** than EU **[22%]** **[*p*<0.001]**. • Firmicutes less in BF **[27%]** than EU **[64%]** **[*p*<0.001]**. • Proteobacteria less in BF **[1%]** than EU **[8%]**. *Genus:* • *Prevotella* and *Xylanibacter* found in BF but not in EU: reported to be related to high fibre intake. *4. Other:* BF have 4 times more propionic **[*p*=0.001]** and butyric acid than EU **[*p*<0.001]**.	*AD:* Kruskal-Wallis [*p-value*] [F/B ratio ±SD] *Taxa:* ANOVA and Kruskal-Wallis [%Relative Abundance and *p-value*] *Other:* SCFAs stat= one-tailed student ttest [*p-value*]	• Diet is related to gut microbiota mostly in an inferred manner. • It is good to note that past 6 mos antibiotic and hospitalization were exclusion criteria for this study. It would be better to use other statistical methods that could take into account other confounding factors that may affect the gut microbiota.

(Reference) **Country** **Study name if applicable**	**Participant age** **Sample size used in microbiota analyses (sex)**	**Study type** **Dietary method** **Gut microbiota method**	**Main findings (Diet with faecal gut microbiota)** 1. ***Alpha diversity, AD*** 2. ***Beta diversity/Clusters, BDC*** 3. ***Taxa; colour code:*** Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria 4. ***Other***	**Stat used and confounders adj for**	**Comments**
(Amarri et al., 2006) Italy	4 – 9 mos 22 (male = 14)	Descriptive (5 mos) 7d dietary record every 1 mo. Microbiological agar plates (culture dependent method)	*3. Taxa* (log CFU per g wet faeces): *Genus:* • *Enterobacteria* **[*p*<0.05]** and *Enterococci* **[*p*=0.02]** increase with age: related to diet modifications during CF had no impact on *Bifidobacteria* but increased those of *Enterobacteria* and *Enterococci.*	*Taxa:* Linear mixed model with age as the fixed effect and infant as the random effect	• Inclusion criteria include EBF for 4 mos, and no antibiotic use 1 mo before study. • Diet captured number of servings of different food groups. However, no analyses were done looking at gut microbiota and diet.

Abbreviations: RCT, randomised controlled trial; FD, Faith's phylogenetic diversity; OTU, operational taxonomic unit; BH, Benjamini-Hochberg; DMM, Dirichlet multinomial mixtures; JSD, Jensen-Shannon divergence; CH, Calinski–Harabasz; GLM, general linear model; CAP, constrained analysis of principal coordinates; db-RDA, distance-based redundancy analysis; PICRUSt, Phylogenetic Investigation of Communities by Reconstruction of Unobserved States; SCFA, short chain fatty acids; CI, confidence interval; -ve, negative; +ve, positive; var, variance; F/B, Firmicutes/Bacteriodetes ratio; adj = , adjusted for; FDR, false discovery rate; stat, statistic; wt, weighted; unwt, unweighted; diff, difference; PCA, principal component analysis; PCoA, principal coordinate analysis; sig, significant; log, logistic; rep, repeated; qPCR; quantitative polymerase chain reaction, also known as real-time PCR; ROS, reactive oxygen species; CFU, colony forming unit; EBF, exclusively breastfed; CHO, carbohydrate; FFQ, food frequency questionnaire; CF, complementary feeding; d, day; mo, month; y, year.

2.11 Associations between diet and gut microbiota in obesity

The focus of this thesis is on effects of the diet on the gut microbiota. However, the reason why both are important is because of their possible association with health outcomes, particularly obesity, in children. Obesity is defined as abnormal or excessive fat accumulation that presents a risk to health. The prevalence of obesity is high in New Zealand, with 1 in 8 children (2 – 14 years) being obese (Ministry of Health, 2019). Obese children are at a higher risk of becoming obese adults, with a review finding 1 in 3 obese pre-school children are also obese as adults (Serdula et al., 1993).

In two studies in children (3 – 18 years and 6 – 16 years), obese children had higher Firmicutes and lower Bacteroidetes compared to normal weight children (Hou et al., 2017; Riva et al., 2017). In addition, obese children had greater amounts of SCFAs (Riva et al., 2017), which is supported by studies in animals which found that obese-type mice had increased faecal SCFA concentrations (Murphy et al., 2010). This observation has been related to increased energy harvesting capacity of the gut microbiota and changes in energy balance (Turnbaugh et al., 2006). However, this is a conflicting result, as SCFAs also have known beneficial effects on the host health in relation to anti-inflammatory properties, strengthening the gut barrier and protection against development of colorectal cancer (Ríos-Covián et al., 2016). A recent study hypothesized that a possible rationale for associations of the higher faecal SCFA concentrations with obesity in some studies, may be due to less efficient absorption and utilization of the SCFAs and hence the SCFAs are found in the faecal samples (de la Cuesta-Zuluaga et al., 2018). To add to the discussion, others have found no difference in the F/B ratio between obese and normal weight adolescents (Hu et al., 2015). Possible reasons for the contrary findings in this case, may be due to the different age ranges studied, and also the cut-offs used to define obesity, BMI $\geq 95^{th}$ percentile in the first two studies in children and BMI $\geq 99^{th}$ percentile in the study in adolescents.

In section 2.10, we have observed some associations between diet and the gut microbiota. However, what it is still unknown in the literature is the direction of the associations between diet, the gut microbiota, and obesity. Are associations between the gut microbiota and obesity actually just reflecting associations between *diet* and obesity? Or is gut microbiota an independent predictor of obesity? Or does obesity lead to dysbiosis of the gut microbiota, independent of diet?

2.12 Literature review conclusions

Despite increased interest in the effects of diet on the gut microbiota, there has been limited research on the impact of complementary feeding and children's diets on the gut microbiota. Most of the studies are observational studies. The few studies that are RCTs, looked at the effect of sources of iron on the gut microbiota. Observational studies only enable us to identify associations and, due to the number of associations tested in the studies, these should be considered as exploratory analyses, with many studies relying on small samples and conducting multiple comparisons in a single study. Hence, study designs and statistical methods that would enable detection of causality are urgently required in this field looking at the impact of diet on the gut microbiota in children.

Very few studies have considered the importance of choosing the right dietary assessment method for the study design, and in particular, ensuring that an FFQ has been validated for the study population and nutrients of interest.

In addition, few of the studies summarised in **Table 2.8** have taken into account the features of the gut microbiota data: that they are compositional, are highly dimensional, and have many zeros; or have included confounders such as parity, mode of delivery, or infant feeding in their statistical methods when examining the effect of diet on the gut microbiota.

Hence, the aim of this thesis is to address the following gaps in the literature on diet and gut microbiota in infants and young children:

1. How is diet during the complementary feeding period associated with children's subsequent gut microbiota composition at 12 months?
2. Can an FFQ measure intake of dietary components thought to influence the composition of the gut microbiota?
3. In what ways is diet associated with the composition of the gut microbiota at 5 years of age?

3 BLISS Methods

This chapter describes the overall methods used in the Baby-Led Introduction to SolidS (BLISS) study and covers in detail the methods that are relevant to this thesis. The other methods are described in more detail in the protocol paper for the BLISS study (Daniels et al., 2015), published main outcomes paper (Taylor et al., 2017b), and other candidates' theses (Daniels, 2017; Williams-Erickson, 2015).

3.1 Study design

The BLISS study was a two-arm randomised controlled trial in infants and their caregivers from Dunedin, NZ. The main aim of the study was to investigate whether a version of the Baby-Led Weaning (BLW) approach to complementary feeding that was modified to address concerns about iron deficiency, growth faltering and choking risk encouraged better self-regulation of energy intake and prevented the development of overweight (Taylor et al., 2017b). The BLISS study has full ethical approval from the Lower South Regional Ethics Committee (LRS 11/09/037) and is registered with the Australian New Zealand Clinical Trials Registry ACTRN12612001133820. Written informed consent was obtained before birth from the primary caregiver of each infant enrolled in the study and before collection of the faecal sample for gut microbiota analysis.

3.1.1 Participants

The BLISS study recruited 206 pregnant women between November 2012 and February 2014 from the Queen Mary Maternity Centre (Dunedin Hospital, Dunedin, NZ). Queen Mary Maternity Centre is the only birthing unit in the city of Dunedin and provides primary, secondary and tertiary maternity services for more than 97% of births in Dunedin city. Inclusion criteria were: mothers who were booked into Queen Mary Maternity Centre before 35 weeks' gestation, intended to live in Dunedin for the next two years, were able to communicate in English or Te Reo Māori (the language of the indigenous people of New Zealand), were 16 years or older, and had a singleton birth. Exclusion criteria were: infants born before 36.5 weeks gestation, and infants having an abnormality or illness detected after birth that was likely to affect their feeding or growth. Eligible families received a study invitation letter and information sheet and an option to opt-out.

3.1.2 Sample size

The power calculation was conducted based on sufficient power to detect a clinically significant difference in body mass index at 12 months, which is the primary outcome of the BLISS study. For the microbiota component, which is the main outcome in this thesis, sample size was not calculated, as it was an additional measure within BLISS. As funding for the microbiota component of the study only became

available in May 2014, faecal samples for microbiota analysis were only collected from infants after this time. Of the 165 participants in the BLISS study at that time, 86 were 7 months of age or younger (and therefore eligible to participate), 74 (86%) of whom agreed to provide a faecal sample for microbiota analyses.

3.1.3 Randomisation

Randomisation was undertaken after the Maternal Baseline Questionnaire was completed during pregnancy. Participants were randomised to the Control (n = 106) or BLISS (n = 108) groups via random length blocks (maximum of 7) after stratification for parity (first child or subsequent child) and maternal education (non-tertiary or tertiary). Allocation was concealed using opaque pre-sealed envelopes. After birth, 8 infants were excluded so the final sample for the BLISS study was Control (n = 101) and BLISS (n = 105). Of the 74 who agreed to provide a faecal sample, the number in each group was equal: Control (n = 37) and BLISS (n = 37) (**Figure 3.1**).

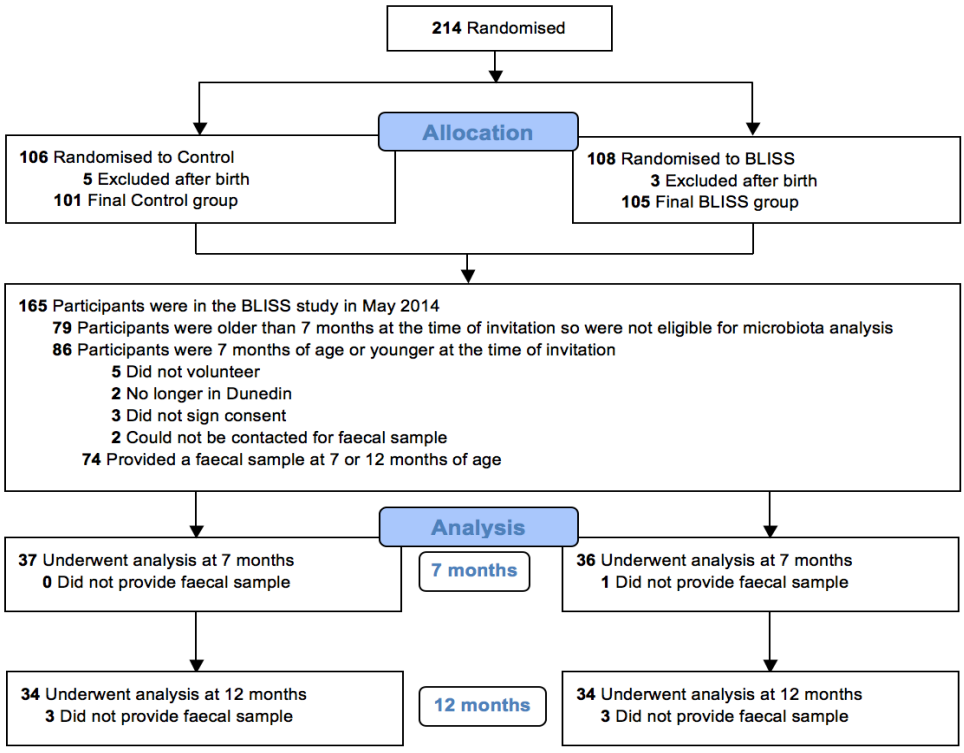


Figure 3.1: CONSORT flow diagram for the microbiota component of the BLISS study, as published in Leong et al. (2018a).

3.2 Study groups and intervention

Both groups received routine midwifery care until 6 weeks of age and Well Child care after this time. Well Child Tamariki Ora is a free health care programme for all children in NZ under 5 years of age. This programme involves free home and clinic visits by trained community nurses who provide advice that covers child growth and development, and checks on oral health, vision, hearing and overall health and development (Ministry of Health, 2017).

3.2.1 Control

The Control group received Well Child Tamariki Ora routine care (as described in section 3.2), with no additional intervention.

3.2.2 BLISS

The BLISS group received Well Child Tamariki Ora routine care (as described in section 3.2), and additional parent contacts from the BLISS study team for support and education from birth to 9 months of age. The BLISS study team that delivered the intervention included: an experienced International Board-Certified Lactation Consultant (IBCLC), and trained research staff who were supervised by a multidisciplinary team (dietitian, paediatrician and speech-language therapist). The intervention had three key components, which were provided at different time points as summarised in **Table 3.1**, namely, contact with an IBCLC, BLISS advice, and BLISS resources.

3.2.3 Adherence

Adherence to a baby-led approach to infant feeding was defined as the infant feeding themselves most or all their food in the past week at 7 and 12 months. This was determined through questionnaires which asked: 'how has your baby been fed solids in the past week?', with responses: 1) fed by an adult, 2) mostly fed by an adult and some baby fed themselves, 3) about half spoon-fed by an adult and half baby fed themselves, 4) mostly baby fed themselves and some fed by an adult, or 5) baby fed themselves. Responses 4 and 5 were determined to be adhering to a baby-led approach to infant feeding.

Table 3.1: Summary of intervention given to BLISS participants.

IBCLC	**BLISS advice**	**BLISS resources**
Third trimester of pregnancy to 6 months	5.5 to 9 months	Third trimester of pregnancy to 9 months
Five or more contacts	Three or more contacts	A range of resources
• An anticipatory guidance group session before birth, which discussed breastfeeding, the free support given and the concept of BLISS (first contact). • Home visits or phone calls from the first week to five months to provide support and encourage exclusive breastfeeding to 6 months, breastfeeding to at least 12 months and delaying the introduction of complementary foods until 6 months (four contacts). • Additional support was also provided when requested by the participant to give specific advice to address individual problems with breastfeeding (or formula feeding) via extra home visit(s), phone or email contact.	• Home visits to provide individualised advice on the introduction of complementary foods using the BLISS approach. • Participants were advised not to start BLISS until their infant was 6 months, to encourage responsive feeding and to offer 'easy' foods and more frequent milk feeds during both illness and recovery. • Parents were encouraged to offer three food types at each meal: 1. An iron-rich food (e.g., red meat, iron-fortified infant cereal). 2. An energy-rich food (>1.5 kcal/g, e.g., avocado, cheese). 3. An easy food (e.g., fruit or vegetable). • Research staff were also available to provide additional support if requested by the participant.	• Resources included information about the BLISS study, recipe books, everyday food lists, and safety information. • The resources encouraged parents to: 1. Test foods before they were offered to ensure they were soft enough to mash with the tongue on the roof of the mouth (or were large and fibrous enough that small pieces did not break off when sucked and chewed, e.g., strips of meat). 2. Avoid offering foods that formed a crumb in the mouth. 3. Offer foods that were at least as long as the child's fist, on at least one side of the food. 4. Ensure the infant was always sitting upright when they were eating – never leaning backwards. 5. Always have an adult with the child when they were eating. 6. Never put whole foods into the infant's mouth – the infant was to do this at their own pace and under their own control.

Abbreviations: IBCLC, International Board-Certified Lactation Consultant; BLISS, Baby-Led Introduction to SolidS.

3.3 Questionnaire data

3.3.1 Maternal baseline questionnaire

Demographic data such as maternal age, ethnicity, education and parity were collected at baseline through the maternal baseline questionnaire. The New Zealand

Index of Deprivation (NZDep) score for each household (Atkinson, Salmond, & Crampton, 2014) was determined using the participant's current address – this index provides a broad indicator of area level deprivation and ranges from 1 (low deprivation) to 10 (high deprivation). Infant sex, birth weight, and gestational age at birth were accessed through hospital records.

3.3.2 Feeding questionnaire

A brief feeding questionnaire was administered at 7 and 12 months to collect information about whether the infant was breast, or formula, fed at what age breastfeeding stopped, and/or the age when formula feeding started and stopped.

3.3.3 Microbiota questionnaire

At 7 and 12 months, a brief microbiota questionnaire (**Appendix A**) was given to participants to complete which included questions on mode of delivery, recent antibiotic usage and yoghurt consumption.

3.4 Dietary assessment

Three-day weighed diet records (WDRs) were used to assess dietary intake at 7 and 12 months. Participants were given detailed written and verbal instructions on how to complete the WDR for their child and then completed three randomly assigned non-consecutive days (2 week days and 1 weekend day) over a three-week period. To control for day-of-the-week effects, each day of the week was represented approximately an equal number of times among participants. Dietary weighing scales (Salter Electronic, Salter Housewares Ltd. Tonbridge, United Kingdom), accurate to ± 1 g were provided to each participant. The diet record (**Appendix B**) contained four components that allowed participants to record:

1. The time of the day, type and brand of the food or drink, preparation method and consistency of the food or drink (puréed, mashed, diced or whole), and the total weight of the food or drink before it was offered to the child and after the child had finished eating so that leftovers could be accounted for.
2. Description of any recipes used.
3. An end of day questionnaire to determine whether it was a typical eating day for the child and how the meals compared to those consumed by the rest of the family.

4. Whether there was any supplement usage.

Once the WDR had been completed and collected, a research staff member would check the WDR to ensure that no information was missing. If any information was missing or unclear, the participant would be contacted for clarification.

3.5 Dietary analysis

The completed WDRs were entered into Kai-culator (Version 1.13s, University of Otago, New Zealand) following a protocol and codebook developed by the BLISS team. Kai-culator is a dietary analysis software programme that includes: dietary data from the New Zealand Food Composition Database (FOODfiles 2010, Plant and Food Research) (New Zealand Institute for Plant & Food Research Limited & Ministry of Health, 2010), recipes for commonly consumed mixed dishes in the 2008/09 New Zealand Adult Nutrition Survey (University of Otago and Ministry of Health, 2011), and commercial infant foods collated by the research team (Clouston, 2014). Once the WDRs had been entered, registered dietitians then checked them for accuracy and consistency.

3.5.1 Infant milk analysis

Breast milk intake was not measured in this study. Instead, total breast milk intake was estimated using total daily volumes of 750 grams per day at 7 months and 448 grams per day at 12 months (Dewey, Heinig, Nommsen, & Lonnerdal, 1991a). These amounts were determined to be the most appropriate estimation of breast milk intakes based on previous literature, at 7 months by a previous Masters candidate (Williams-Erickson, 2015), and at 12 months by a previous PhD candidate (Daniels, 2017). The amount of infant formula consumed was determined by the amount that the participants recorded in the WDR. If the child was mixed fed (fed both breast milk and infant formula), the amount of breast milk consumed by the child could be calculated as the estimated total breast milk intake (i.e. 750 g or 448 g per day) minus the amount of infant formula consumed.

3.5.2 Nutrient analysis

Data on the key nutrients and food components of interest (energy and dietary fibre) were sourced from Kai-culator. The values for amount consumed per day were exported from Kai-culator to be used in the statistical analysis.

3.5.3 Food group analysis

The Candidate developed 18 food groups that were considered, based on the literature, to have the potential to be relevant to the gut microbiota (**Table 3.2**). The food groups were modified from two previous theses that had been designed to investigate overall diet at 7 months (Williams-Erickson, 2015), and iron and zinc intake (Daniels, 2017). In total, there were 1682 individual food items that had been consumed by the children in the WDRs at 7 and 12 months. Each of these individual foods was assigned to one of the 18 food groups that had been developed (**Table 3.2**). An Excel spreadsheet was then created with the energy (kJ) per 100 g for each of the 1682 individual food items (**Appendix C**). For each food item it was necessary to determine the contribution of that food item to the energy provided from each of the 18 food groups (per 100 g). The Candidate reassigned 200 of the individual food items so that they were in the appropriate microbiota-relevant food group. The Candidate then carried out technical quality control checks on the data for both 7 and 12 month WDRs.

3.5.4 Fibre variety score

A dietary fibre variety score was determined by counting each different fibre containing food (i.e. grain product, vegetable or fruit) consumed over the 3 recording days. Hence, only those records that had 3 complete days were counted. Mixed food dishes and commercial baby foods were broken down into their component parts to be counted. Condiments were not included in the variety score count. Variety scores were calculated using a method similar to that used in a study looking at food variety at 2 years of age (Scott, Chih, & Oddy, 2012) as there were no reported methods specifically for 7 and 12 month old infants. Tabulation of the dietary fibre variety score was carried out and reported in a previous candidate's thesis (Morison, 2015).

Table 3.2: Individual food items included in each food group for food group analysis.

Food Group (Acronym)	Foods Included
1. Breast milk (BM)	Breast milk
2. Infant formula (IFO)	All infant formulas
3. Cow's milk (CM)	Cow's milk (including milk powder)
4. Dairy (DAI)	Cheese, yoghurt, sour cream, custard, cream, flavoured milk, dairy food
5. Fruit and fruit juice (FF)	Apple, banana, orange, mandarin, kiwifruit, mango, melon, grapes, pears, apple sauce, berries (strawberries, blueberries etc.), avocado, fruit juice, other fruit - fresh, canned, stewed, poached, dried, juiced
6. Vegetables (VEG)	Potato (mashed, baked etc.), kumara, pumpkin, carrot, green beans, peas, corn, capsicum, broccoli, cauliflower, courgette, mushroom, green leafy vegetables, mixed vegetables, tomatoes (fresh, canned, pureed, paste), olives, vegetable soup, other vegetables - frozen, fresh, canned, in brine
7. Legumes (LEG)	Beans (lentils, chickpeas, kidney beans, baked beans), hummus, tofu, tempeh, sausages (vegetarian)
8. Nuts and seeds (NUT)	All nuts and seeds, nut butters, tahini
9. Iron fortified infant cereal (IIC)	Baby rice cereal, baby muesli, baby porridge, iron-fortified teething rusks
10. Breads and cereals (BC)	Plain crackers (cruskits, corn crispbread, cream crackers, rice crackers, wholemeal crackers, rice cakes, cheese crackers etc.), English muffins, bread buns, bread (white, wholemeal, multigrain, rye etc.), pizza bases, cereals (weetbix, cornflakes, rice bubbles, porridge, oats, Special K, Sultana bran etc.), noodles (udon, egg, rice, instant, chow mien etc.), rice, pasta, couscous, quinoa, chia, rusks, spaghetti in tomato sauce, crumpet
11. Eggs (EGG)	Eggs (boiled, scrambled, poached etc.)
12. Red meat (RM)	Beef, lamb, mutton, veal, venison, red meat offal (beef kidney, lamb liver)
13. Fish and poultry (FP)	Fish (all types), poultry
14. Other meat (OM)	Pork, processed meat (bacon, ham, salami, sausages, pastrami), chicken liver pate, meatloaf
15. Sweet food (SWE)	Scones (plain), waffles, baked goods (biscuits, cookies, cake, muffins, loaves, iced buns, sweet pastries, waffle cones, crepe), pikelets, pancakes, rice pudding, tapioca pudding, ice cream, sweets (jelly, lollies, ice blocks, sorbet, chocolate, maple syrup), muesli bars
16. Savoury food (SAV)	Scones (cheese), popcorn, pastry sheets, baked goods (croissants, savoury pastries), sausage rolls, cheese rolls, meat pies, fried foods (fries/potato chips, hot chips, fritters, McDonalds, KFC), corn chips, extruded corn snacks
17. Beverages (BEV)	Rice milk, soy milk, almond milk, coconut milk, raro juice, drinking chocolate, smoothies, tea, milo, coffee
18. Miscellaneous (MIS)	All other foods, including: fats (butter, oil, margarine), sauces and stocks (mayonnaise, dressings, gravy, white sauce, tomato sauce and relish, pesto, coconut cream, curry paste, pizza sauce), spreads (marmite, vegemite, jam, cheese spread, nutella, cocoa etc.), condiments (spices, herbs, vinegar, seaweed, desiccated coconut)

3.6 Gut microbiota

3.6.1 Faecal collection

Participants were asked to collect a faecal sample (approximately the size of a New Zealand 20 cent coin which is approximately 22 mm in diameter, using disposable gloves, and avoiding urine contamination) at 7 and 12 months of age from their child's nappy. A specimen collection instruction sheet was given to the participants, which gave detailed instructions on how the faecal sample was to be collected (**Appendix D**). The participants were given a specimen jar with a scoop, container, plastic bag, disposable gloves and an insulated flask. The sample was then stored in the home freezer (-18°C) in study-provided freezer containers before collection and delivery to the Department of Microbiology and Immunology, University of Otago, where they were stored at -80°C until DNA extraction.

3.6.2 Data preparation

DNA was extracted from 250 mg faeces according to the protocol provided by the manufacturer (12855-100 PowerSoil DNA Isolation Kit; Mo Bio). The MoBio kit uses both mechanical and chemical lysis methods. Amplification of the 16S rRNA gene V4 region, library preparation and sequencing were carried out at Argonne National Laboratories (University of Chicago) using 2 x 250 base paired-end reads on an Illumina MiSeq instrument. The data preparation was carried out by Blair Lawley from the Department of Microbiology and Immunology, University of Otago, and can be summarised in the following steps:

1. Join paired ends - the sequencing was carried out as two reads, one from each end of the V4 amplicon. Hence for each Polymerase Chain Reaction (PCR) product, two sequences were read: one from the forward primer end and another from the reverse primer end. The two reads were combined to generate a single read.
2. Split libraries - the joined reads were checked for quality and associated with sample numbers. This was achieved via the unique barcode attached to amplicons from each sample. Poor quality reads were discarded at this point and reads were re-named and numbered according to the sample that they came from.
3. Sequence de-replication - the sequences were compared, and any identical sequences were clustered together at 100% identity. The number of sequences

belonging to a cluster was stored and a single sequence from each cluster was used for downstream work.

4. Clustering at 97% similarity - the de-replicated sequences were then clustered using a 97% similarity cut off point. This step removed clusters that contain only one sequence (known as 'singletons'), which were likely to be sequence artefacts.
5. Chimera removal - chimeras were formed when amplifying 16S rRNA gene targets from complex communities. Denovo (independent of a database, using clustering numbers) and reference (comparing sequences to a database) approaches were used to detect the chimeras. The chimeras were removed to reduce noise.
6. Map original sequences onto filtered and clustered sequences - the filtered and clustered sequence set was used as a 'database' to map all original sequences. Hence, all the sequences available were used and poor-quality sequences were placed into sensible, quality-controlled clusters. This helped to improve data retention while maintaining sequence quality.
7. Generate Operational Taxonomic Unit (OTU) table - a table was created with samples as column headers and OTUs as row headers. It was used for all downstream diversity and comparison work.
8. Add taxonomy - representative sequences from each OTU were compared with a database. The top 3 to 5 best matches with the database were returned and taxonomy was assigned. If the top matches disagreed for taxonomy assignment, the taxonomy was taken at the highest level at which agreement was reached. Hence, some OTUs were classified to genus level while others may only be classified to family level or higher.

The Sequences were analysed using a combination of QIIME version 1.9.1 (Caporaso et al., 2010) and vsearch version 1.9.5 (Rognes et al., 2016). Taxonomy classifications were made using the SILVA version 123 database (Quast et al., 2013). Raw sequence data can be publicly assessed at NCBI (Accession Numbers: SAMN08045869 & SAMN08045871).

3.7 Statistical analysis

All relevant statistical analyses are discussed in **Chapter 4**.

4 A modified version of baby-led weaning on the infant gut microbiota

This chapter investigates the effect of a modified version of baby-led weaning (BLW) on the infant gut microbiota and uses mediation analysis to identify characteristics of the diet that may impact the infant gut microbiota. The statistical analysis plan and lessons learnt are included in **Appendix E**.

This chapter uses the following data collected during the BLISS study:

- Maternal Baseline Questionnaire (collected during pregnancy).
- Feeding Questionnaires (collected at 7 and 12 months).
- Microbiota Questionnaire (collected at 7 and 12 months).
- Gut microbiota data from faecal samples (collected at 7 and 12 months).
- Three-day weighed diet records (WDR) (collected at 7 and 12 months).

Chapter highlights:

- Use of mediation analyses in gut microbiota analyses. This approach is commonly used in the health literature, but to the Candidate's knowledge, this is the first study to use mediation analysis to determine the pathway of the relationship between a predictor and the gut microbiota.
- How a baby-led approach to complementary feeding may impact the gut microbiota, which has not been examined previously.
- High quality dietary data in infants using 3-day WDRs, which is not commonly found in literature investigating diet and the gut microbiota. Some studies use 24-hour recalls or FFQs that have not been validated for that study population.

A paper based on this chapter has been published:

Leong, C., Haszard, J. J., Lawley, B., Otal, A., Taylor, R. W., Szymlek-Gay, E. A., Fleming, E. A., Daniels, L., Fangupo, L. J., Tannock, G. W., Heath, A.-L. M. (2018). Mediation analysis as a means of identifying dietary components that affect the fecal microbiota of infants weaned by modified baby-led, compared to traditional, approaches. *Applied and Environmental Microbiology, 84*(18), e00914-00918. doi:10.1128/aem.00914-18

4.1 Introduction

The introduction of complementary foods following the milk-only diet of early life (known in some countries as 'weaning') results in significant changes to the composition of the gut microbiota (Bergstrom et al., 2014; Fallani et al., 2011). Further major changes happen with the introduction of family foods, usually late in the first year of life (Laursen et al., 2016). Traditionally, parents have been encouraged to start spoon-feeding their infant puréed foods from around 6 months of age, progressing to mashed, then chopped foods in the hope that they will be eating family foods by around 12 months (Ministry of Health, 2008). However, an alternative method of complementary feeding, known as BLW, is becoming popular in New Zealand (Morison et al., 2016), the United Kingdom (Brown & Lee, 2011), the United States (Beal, 2016) and Canada (D'Andrea et al., 2016).

In BLW, infants are encouraged to feed themselves whole pieces of food from the family meal from 6 months of age, instead of being offered 'baby food' (Rapley & Murkett, 2008). As a result, infants following BLW are more likely to eat the same foods as the rest of the family than those who are traditionally spoon-fed (Morison et al., 2016). The Baby-Led Introduction to SolidS (BLISS) randomised controlled trial (Cameron, Taylor, & Heath, 2015; Daniels et al., 2015) investigated the impact of a modified version of BLW on growth (Taylor et al., 2017b), choking risk (Fangupo et al., 2016), and iron status (Daniels et al., 2018). In this Chapter, the BLISS study provides an opportunity to investigate the impact that this alternative method of complementary feeding has on the developing infant gut microbiota.

Considerable attention has been paid to the impact of breast milk and infant formula on the developing gut microbiota, with studies finding that infants who were formula-fed have increased *Clostridium difficile* compared to breastfed infants (Azad et al., 2013; Penders et al., 2006), less *Bifidobacteria* (Bezirtzoglou, Tsiotsias, & Welling, 2011), and increased alpha diversity (Ho et al., 2018). Ceasing breastfeeding appears to have a greater impact on maturation of the infant gut microbiota than the timing of introduction of solid foods (Bäckhed et al., 2015; Bergstrom et al., 2014). However, little research has investigated the impact of the change in diet that occurs when infants are introduced to solid foods during the complementary feeding period (Laursen et al., 2017). To date, the majority of studies on this transition have been observational

(Bergstrom et al., 2014; Fallani et al., 2011), and the few randomised controlled trials (RCTs) have focused largely on the impact of iron fortification on the gut microbiota (Krebs et al., 2013; Qasem et al., 2017). Moreover, the time point of 7 months was examined as this would give the infants sufficient time (around 1 month for exclusively breastfed infants) to start complementary food. In addition, we looked at the faecal sample again at 12 months as this was the period where most of the infants would be expected to be eating family foods.

Hence, the BLISS RCT study provides a good source of data to study the infant gut microbiota during this complementary feeding period. We are not only able to look at the method of feeding, but also the diet and how this may affect the infant gut microbiota. The objectives addressed in this chapter were to determine:

1. the gut microbial composition of infants at 7 and 12 months following BLISS compared with infants following traditional spoon-feeding; and
2. the role of diet in any gut microbial differences found.

4.2 Methods

4.2.1 Data collection

Please refer to **Chapter 3** for details on the methods used for the BLISS study. In short, the BLISS study was a 2-year randomised controlled trial that compared a version of BLW (i.e. the infants feed themselves family foods, ideally while breastfeeding on demand) modified to address concerns about choking, iron deficiency and growth faltering (Cameron, Heath, & Taylor, 2012b) with a Control group (receiving usual care and therefore predominantly traditional spoon-feeding). A sub-set of the BLISS study participants provided faecal samples at 2 time points, 7 months and 12 months, that could be used in this thesis: at 7 months, Control (n = 37) and BLISS (n = 36); at 12 months, Control (n = 34) and BLISS (n = 34). There were no significant differences in maternal age, BMI, parity, education, household deprivation, the sex or birth weight of the child between participants who did (n = 74) and did not (n = 132) provide a faecal sample. Both groups received routine midwifery care until 6 weeks of age and Well Child care from then until the end of the study; and BLISS participants received eight additional visits (from before birth to 9 months) which provided support, education and resources on following the BLISS approach. WDRs (section 3.4) and

faecal samples for gut microbiota analysis (section 3.6) were collected at 7 and 12 months of age.

4.2.2 Data preparation

In this Chapter, microbiota composition was described using five alpha diversity measures: number of Operational Taxonomic Units (OTUs; a proxy for observed species), phylogenetic diversity, Chao1 estimator, Simpson Index, and Shannon Index. Therefore, three indices described microbial richness alone (i.e. number of species): observed species, phylogenetic diversity, and Chao1 estimator; and two indices described richness and evenness (i.e. the equality of distribution of the species' frequencies): Simpson Index, and Shannon Index. The OTU table was rarefied to the minimum sample count (30,000 sequences) (**Figure 4.1**) for calculation of the alpha diversity measures.

Alpha diversity was the outcome measure used to describe the gut microbiota in this chapter because a species-rich gut community is less susceptible to foreign invasion. From many studies in adults, alpha diversity plays an important role as a description of the gut microbiota, from association with obesity (Le Chatelier et al., 2013), weight gain (Menni et al., 2017), inflammatory bowel disease (IBD) (Ni, Wu, Albenberg, & Tomov, 2017; Walters, Xu, & Knight, 2014), to recurrent *Clostridium difficile*- associated disease (CDAD) (Lozupone, Stombaugh, Gordon, Jansson, & Knight, 2012).

In addition to alpha diversity measures, relative abundance at the family level was calculated by collapsing the raw OTU table based on seven-level taxonomy strings (kingdom_phylum_class_order_family_genus_species) obtained from the SILVA version 123 database.

For the dietary data, 9 food groups were defined based on food groups that are of relevance to the gut microbiota, and the number of consumers (i.e. it was decided that at least 8 consumers were required in each food group so that there was sufficient power to perform the food group analyses): 'breast milk', 'infant formula', 'dairy', 'fruit, vegetables, nuts and legumes', 'breads and cereals', 'meat, fish and poultry', 'sweet food', 'savoury food' and 'miscellaneous'. **Table 4.1** shows the relationship between the 18 original food groups (**Table 3.2**) and the 9 food groups analysed in this chapter. Ingredients in recipes were coded into their specific food groups instead of the

whole recipe being assigned to a diverse 'mixed dishes' category to enable a more accurate representation of the food groups. For example, if a ham sandwich was eaten, the bread would be under 'breads and cereals', ham under 'meat, fish poultry' and tomato under 'fruit, vegetables, nuts and legumes'.

Table 4.1: Food groups used for analysis in this Chapter.

Original 18 food groups [a]	**9 food groups used in this Chapter**	
Breads and cereals (BC)	Breast milk	BM
Cow's milk (CM)	Infant formula	IFO
Iron fortified infant cereal (IIC)	Dairy	DAI + CM
Fruit and fruit juice (FF)	Fruit, vegetables, nuts and legumes	FF + VEG + NUT + LEG
Vegetables (VEG)		
Dairy (DAI)	Breads and cereals	BC + IIC
Legumes (LEG)	Meat, fish and poultry	EGG + RM + FP + OM
Nuts and seeds (NUT)	Sweet food	SWE
Eggs (EGG)	Savoury food	SAV
Red meat (RM)	Miscellaneous	MIS + BEV
Fish and poultry (FP)		
Other meat (OM)		
Beverages (BEV)		
Miscellaneous (MIS)		
Breast milk (BM)		
Infant formula (IFO)		
Sweet food (SWE)		
Savoury food (SAV)		

[a] Created in **Chapter 3**, Table 3.2.

4.2.3 Statistical analysis

Data were analysed using Stata software (StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP). **Table 4.2** shows the baseline measurements of both Control and BLISS groups. P-values were not calculated for baseline measurements as the groups were randomised.

A chi-squared test was used to compare infant formula consumption at 7 months, and a log-rank test (i.e. time to events) to compare the age of introduction of formula, solids, and infant cereal, between study groups (**Table 4.3**). To determine differences in alpha diversity between study groups at both 7 and 12 months of age, linear regression was used with adjustment for parity and maternal education (i.e. the stratification variables within the randomised controlled trial). Adjusted mean differences between the study groups, 95% Confidence Intervals (CI), and p-values were calculated along

with the proportion of variance explained (R^2) by the independent variables (**Figure 4.1** and **Table 4.4**). Subsequent analyses for alpha diversity were then limited to one measure that described richness only, and one measure that described richness and evenness. The two measures chosen (number of OTUs and Shannon Index) had the highest R^2 (i.e. explained the highest proportion of the variance) and had the strongest association with study group (**Table 4.4**).

Assessment of study group differences in relative abundance of the most abundant bacterial families (i.e. families for which median relative abundance was greater than 1%) was determined using median regression (**Figure 4.2** and **Table 4.5**). Pictorial representations of the most abundant bacterial families between study groups is shown in **Figure 4.2**. In **Figure 4.3**, a pictorial representation of the food group intake between study groups is indicated.

The data from both Control and BLISS groups were then combined for cross-sectional analyses. A mixed regression model was used to assess age differences in relative abundance of the most abundant bacterial families, which included participant identification number as a random effect (**Table 4.6**). Linear regression was used to assess whether demographic, feeding variables and antibiotic use in the week before faecal sampling, were related to alpha diversity, with adjustment for study group (**Table 4.7**). The demographic and other variables were chosen based on previous research as explained in the statistical analysis plan (**Appendix E**), and whether the variable was collected in the study. Thereafter, the factor that was found to be significantly associated with alpha diversity was considered a confounder and adjusted for in subsequent analyses (parity was found to be a confounder, **Table 4.7**). To explore whether dietary components at 7 months predicted alpha diversity at 12 months, regression models were generated, adjusted for randomised group, maternal education and parity (**Table 4.8**).

Mediation analysis was then used to determine the extent to which the association between study group and alpha-diversity at 12 months was due to differences in food group intake at 7 months, i.e. whether food group intake 'mediated' the association (**Figure 4.5**). The requirements for mediation were considered to exist (Fairchild & McDaniel, 2017) if randomised group predicted alpha diversity at 12 months, and food group intake at 7 months was both related to alpha diversity at 12 months and different

between the randomised groups. The 'proportion mediated' was the proportion of the effect size without the mediator ('total effect') that was reduced when the mediator was included in the regression ('direct effect') (**Figure 4.4**). As the sample size was limited, decisions and explanations for mediation analysis are based on both effect sizes and p-values, not solely on p-values. Residuals for all models were plotted and visually assessed for homogeneity of variance, and normality.

4.3 Results

4.3.1 Study population

Faecal samples were obtained from 73 participants at 7 months (1 BLISS missing) and 68 participants at 12 months (3 Control and 3 BLISS missing). The baseline and early feeding characteristics of these participants are shown in **Table 4.2** and **Table 4.3**.

Table 4.2: Characteristics of the BLISS study participants who agreed to provide a faecal sample, as published in Leong et al. (2018a).

Variable [a]	**Control (n = 37)**	**BLISS (n = 37)**
Male, n (%)	19 (51.4)	15 (40.5)
Birth weight, (g)	3443 (531)	3507 (474)
Parity, n (%)		
1	15 (40.5)	14 (37.8)
>1	22 (59.5)	23 (62.2)
Mode of delivery, n (%)		
Vaginal	27 (73.0)	29 (78.4)
Caesarean section	10 (27.0)	8 (21.6)
Maternal age, (years)	31.6 (6.0)	31.7 (4.6)
Maternal self-reported BMI, (kg/m^2)	25.3 (5.9)	25.8 (6.2)
Maternal education, n (%)		
School or post school only [b]	15 (40.5)	21 (56.8)
University	22 (59.5)	16 (43.2)
Household deprivation decile [c], n (%)		
1-3 (low)	11 (29.7)	11 (29.7)
4-7	14 (37.8)	21 (56.8)
8-10 (high)	12 (32.4)	5 (13.5)
Provided dietary data at 7 months, n (%)	34 (92)	32 (86)
Provided dietary data at 12 months, n (%)	29 (78)	29 (78)

Abbreviations: BLISS, Baby-Led Introduction to SolidS; BMI, body mass index (calculated as weight in kilograms divided by height in meters squared). [a] Data presented as mean (SD) unless otherwise indicated. Data were missing for 2 participants for birth weight and 3 for maternal self-reported BMI. p-values not calculated for baseline measurements as the groups are randomised. [b] School or post school only: primary school, secondary school, trade, certificates and diplomas. [c] Determined using the New Zealand Index of Deprivation 2013 (Atkinson et al., 2014). The Index combines 9 variables from the 2013 New Zealand National Census to provide a deprivation score for each meshblock (a geographical unit defined by Statistics New Zealand

that contains approximately 81 people). The score reflects the extent of material and social deprivation and is used to construct deciles from 1 (low deprivation) to 10 (high deprivation).

Table 4.3: Early diet characteristics of the BLISS study participants who agreed to provide a faecal sample, as published in Leong et al. (2018a).

Variable [a]	**Control (n = 37)**	**BLISS (n = 37)**	**p-value**
Consumed infant formula by 7 months, n (%)	20 (54)	18 (50)	0.729 [b]
Age at introduction in weeks, median (25^{th}, 75^{th})			
Infant formula	4.0 (0.6, 26)	4.5 (0.4, 28.2)	0.454 [c]
Any solids	22.8 (21.7, 24.9)	26.0 (23.8, 26)	**0.002** [c]
Infant cereal	23.8 (22.1, 25.5)	26.0 (23.8, 26.4)	**0.032** [c]

Bold = significant ($p < 0.05$). Abbreviations: BLISS, Baby-Led Introduction to SolidS. [a] Participants who provided data for age of introduction of infant formula: n = 45; all solids and infant cereal: n = 62. [b] Chi-square test. [c] Log-rank test.

Infants were predominantly born by vaginal delivery to university-educated mothers. BLISS infants were introduced to complementary foods later than Control infants. No differences were observed in the rates of breastfeeding, or the amounts of breast milk or infant formula consumed, at 7 or 12 months of age. Recent antibiotic use was similar in Control and BLISS infants at 7 (16.7% vs 13.5% $p = 0.707$) and 12 (16.1% vs 8.8% $p = 0.371$) months of age. BLISS infants had very good adherence to the baby-led approach and were significantly ($p < 0.001$) more likely to feed themselves most or all their food than Control infants at every age (for example at 7 months BLISS 74% compared to Control 21%) (Taylor et al., 2017b). Using population-averaged generalized estimating equations with the full sample, BLISS infants were also significantly more likely to be eating the same foods as the rest of their family by 7 months of age; having 3.3 (95% CI: 2.0, 5.6; $p < 0.001$) times the odds of eating the same evening meal as the rest of the family compared to the Control infants (Williams-Erickson et al., 2018).

4.3.2 Gut microbiota

Figure 4.1 shows that alpha diversity, as measured by observed 'species', increased in both groups from 7 to 12 months of age (mean change (95% CI) = 77 (64, 90); $p < 0.001$ from paired t-test). From **Table 4.4**, no significant group differences

were observed in the five measures of alpha diversity at 7 months (all $p > 0.1$) and the proportion of variance explained by group, parity, and maternal university education, was low for observed species, phylogenetic diversity, and Chao1 estimator (all < 2%). For those measures of alpha diversity that included evenness (Simpson Index and Shannon Index) the proportion of variance explained was slightly higher (6.3% and 8.1%, respectively), although still low.

At 12 months of age, however, the BLISS group had significantly lower alpha diversity than the Control group. The BLISS group had a mean of 31 fewer OTUs ($p = 0.028$) and 0.52 lower Shannon Index ($p = 0.006$) than the Control group. The proportion of variance explained at 12 months (between 16% and 21%) was higher than at 7 months (between 0.4% and 8%) (**Table 4.4**). As observed species and Shannon Index were the alpha diversity measures that covered richness and richness-and-evenness with the highest R^2, and had the strongest associations with study group, subsequent analyses of alpha diversity were limited to these two measures.

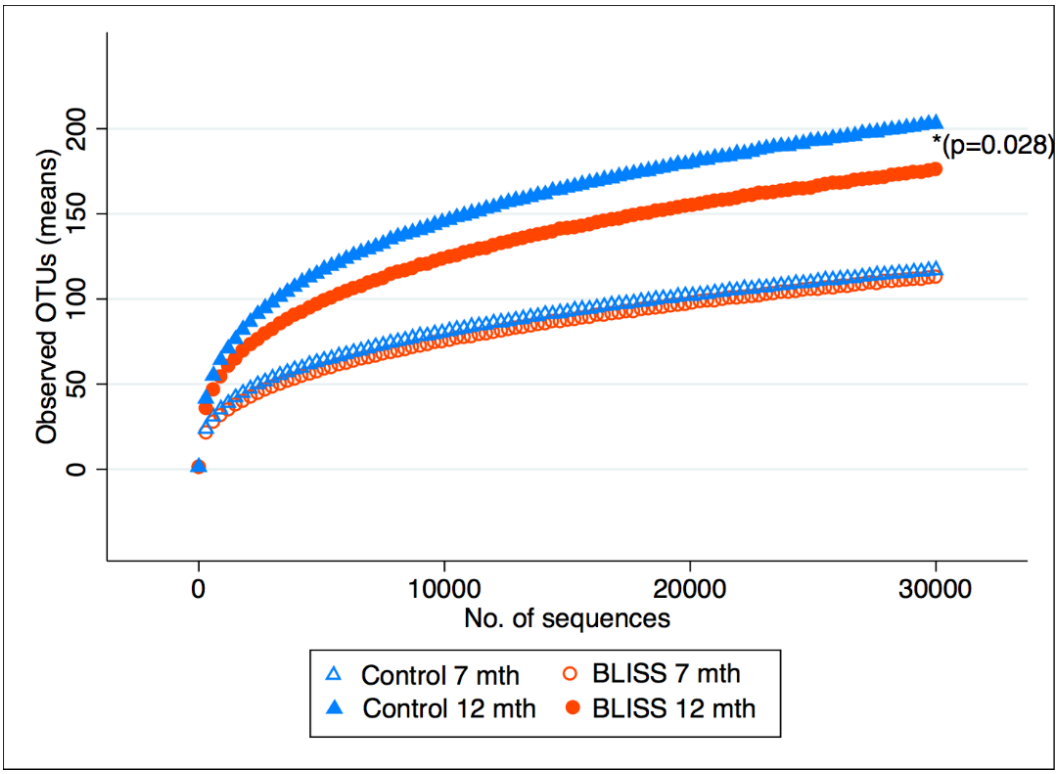


Figure 4.1: Rarefaction curves of observed OTUs against number of sequences from Control and BLISS groups at 7 and 12 months of age, as published in Leong et al. (2018a).

Abbreviations: OTUs, Operational Taxonomic Units (a proxy for observed species; and a measure of alpha diversity); *, p-value between groups at 12 months.

Table 4.4: Alpha diversity measures for Control and BLISS groups at 7 and 12 months, as published in Leong et al. (2018a).

Alpha diversity measures [a]	**Control mean (SD)**	**BLISS mean (SD)**	**Mean difference** [b] **between groups (95% CI)**	**p-value**	**Proportion of variance explained (R^2)**
7 months	(n = 37)	(n = 36)			
Observed species (no. of OTUs)	117 (36)	113 (30)	-5.2 (-21.2, 10.9)	0.523	0.018
Phylogenetic diversity	10.7 (2.0)	10.7 (1.8)	0.04 (-0.88, 0.96)	0.925	0.004
Chao1	174 (46)	173 (39)	-1.6 (-22.0, 18.8)	0.877	0.012
Simpson	0.76 (0.11)	0.74 (0.09)	-0.03 (-0.08, 0.02)	0.187	0.063
Shannon	2.99 (0.72)	2.79 (0.58)	-0.25 (-0.56, 0.06)	0.109	0.081
12 months	(n = 34)	(n = 34)			
Observed species (no. of OTUs)	203 (58)	176 (58)	-30.9 (-58.5, -3.4)	**0.028**	0.171
Phylogenetic diversity	14.7 (2.8)	13.8 (3.3)	-1.2 (-2.6, 0.2)	0.094	0.175
Chao1	277 (69)	246 (69)	-36.4 (-69.1, -3.8)	**0.029**	0.164
Simpson	0.88 (0.06)	0.83 (0.09)	-0.06 (-0.09, -0.02)	**0.004**	0.185
Shannon	4.18 (0.73)	3.72 (0.81)	-0.52 (-0.88, -0.16)	**0.006**	0.210

Bold = significant ($p < 0.05$). Abbreviations: BLISS, Baby-Led Introduction to SolidS; OTUs, Operational Taxonomic Units. [a] Estimates were made at 30,000 sequences. [b] Linear regression adjusted for parity and maternal university education.

Sensitivity analyses were undertaken with the inclusion of antibiotic use in the days before the faecal sample was collected (n = 11 at 7 months; n = 8 at 12 months). No appreciable impact on the 7-month results was seen, but the 12-month results were slightly strengthened by the inclusion of antibiotic use into the model (mean difference (95% CI): observed species -33 (-62, -5) OTUs, R2 = 0.19; Shannon Index -0.56 (-0.94, -0.19), R2 = 0.23).

Figure 4.2 illustrates the relative abundance of the 7 most abundant families (median of at least 1% relative abundance), which in combination explained approximately 86% of normalized sequence data. There were no significant differences in relative abundance of the 7 bacterial families between groups at either 7 or 12 months of age (**Table 4.5**). However, for all infants combined, there was a decrease in the relative abundance of Bifidobacteriaceae, Enterobacteriaceae and Veillonellaceae, with increases in Lachnospiraceae and Ruminococcaceae between 7 and 12 months of age (**Table 4.6**).

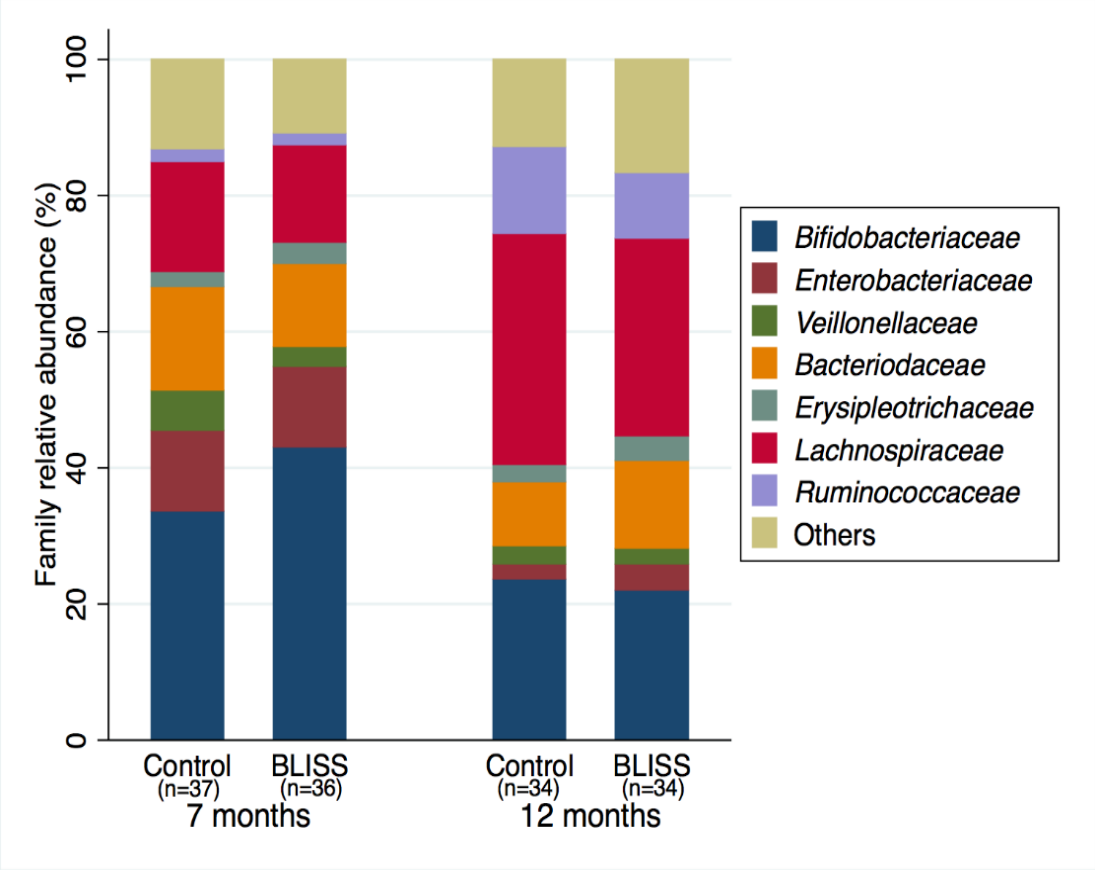


Figure 4.2: Relative abundances of top (>1%) seven bacterial families in faeces of Control and BLISS infants at 7 and 12 months of age, as published in Leong et al. (2018a).

Table 4.5: Median (25th, 75th percentile) relative abundance of the seven most abundant bacterial families according to group.

Family	Control	BLISS	Mean difference (95% CI) [a]	p-value
Bifidobacteriaceae				
7 months	38 (9,45)	44 (38,55)	9.9 (-3.9,24)	0.156
12 months	22 (12,32)	20 (8,30)	0.2 (-9.0,9.4)	0.966
Enterobacteriaceae				
7 months	10 (3,20)	6 (4,20)	-3.0 (-12,5.9)	0.502
12 months	1 (0,2)	0 (0,3)	-0.0 (-0.9,0.9)	0.948
Veillonellaceae				
7 months	4 (1,6)	1 (0,4)	-0.9 (-0.3,1.4)	0.445
12 months	1 (0,3)	1 (1,3)	0.2 (-1.1,1.5)	0.775
Bacteroidaceae				
7 months	12 (0,24)	4 (0,19)	-3.8 (-16,8.6)	0.546
12 months	8 (1,14)	8 (0,20)	-0.0 (-9.3,9.3)	0.999
Erysipelotrichaceae				
7 months	1 (0,2)	1 (0,4)	0.5 (-1.0,1.9)	0.500
12 months	2 (1,3)	2 (1,4)	0.6 (-0.6,1.7)	0.345
Lachnospiraceae				
7 months	11 (1,26)	7 (1,25)	-0.1 (-12,12)	0.981
12 months	35 (25,42)	29 (14,40)	-8.3 (-19,2.2)	0.119
Ruminococcaceae				
7 months	0 (0,1)	0 (0,1)	-0.1 (-0.4,0.6)	0.786
12 months	12 (4,21)	10 (1,16)	-4.9 (-13,3.4)	0.240

Abbreviations: BLISS, Baby-Led Introduction to SolidS. [a] Mean difference using median regression.

Table 4.6: Median (25th, 75th percentile) relative abundance of the seven most abundant bacterial families at 7 and 12 months and the mean difference with time.

Family	7 months (n=73)	12 months (n=68)	Mean difference (95% CI) [a] (n=141)	p-value
Bifidobacteriaceae [b]	42 (24,55)	21 (11,31)	-16 (-21,-10)	**<0.001**
Enterobacteriaceae	7.2 (2.9,20)	0.6 (0.2,2.5)	-8.9 (-12,-5.8)	**<0.001**
Veillonellaceae	2.2 (0.5,5.2)	1.1 (0.6,3.1)	-1.9 (-3.4,-0.3)	**0.020**
Bacteroidaceae	6.0 (0.0,22)	8.0 (1.0,19)	-2.5 (-6.9,1.9)	0.272
Erysipelotrichaceae	0.9 (0.1,3.2)	1.9 (1.1,3.3)	0.4 (-1.0,1.7)	0.586
Lachnospiraceae	9.0 (1.5,26)	32 (19,41)	16 (12,21)	**<0.001**
Ruminococcaceae	0.1 (0.0,0.9)	10 (1.6,19)	9.4 (7.3,12)	**<0.001**

Bold = significant ($p < 0.05$). [a] Mean difference using mixed regression model with participant identification number as a random effect to account for repeated measures. [b] p-for-interaction between group and time was close to significance at $p = 0.050$, with the BLISS group having a greater mean decrease (95% CI) (-21 (-29, -13)) compared to the Control group (-11 (-18, -3.3)).

4.3.3 Diet

Energy intakes were similar in Control and BLISS participants at 7 months of age (mean (SD): 2820 (470) kJ/d and 2860 (440) kJ/d, respectively; p = 0.749) and at 12 months of age (mean (SD): 3460 (850) kJ/d and 3430 (480) kJ/d, respectively; p = 0.881). However, the proportion of energy contributed by each food group changed considerably over time (**Figure 4.3**). Milk (breast milk and infant formula) contributed approximately 78% of energy for both groups at 7 months of age, decreasing to around 36% at 12 months of age.

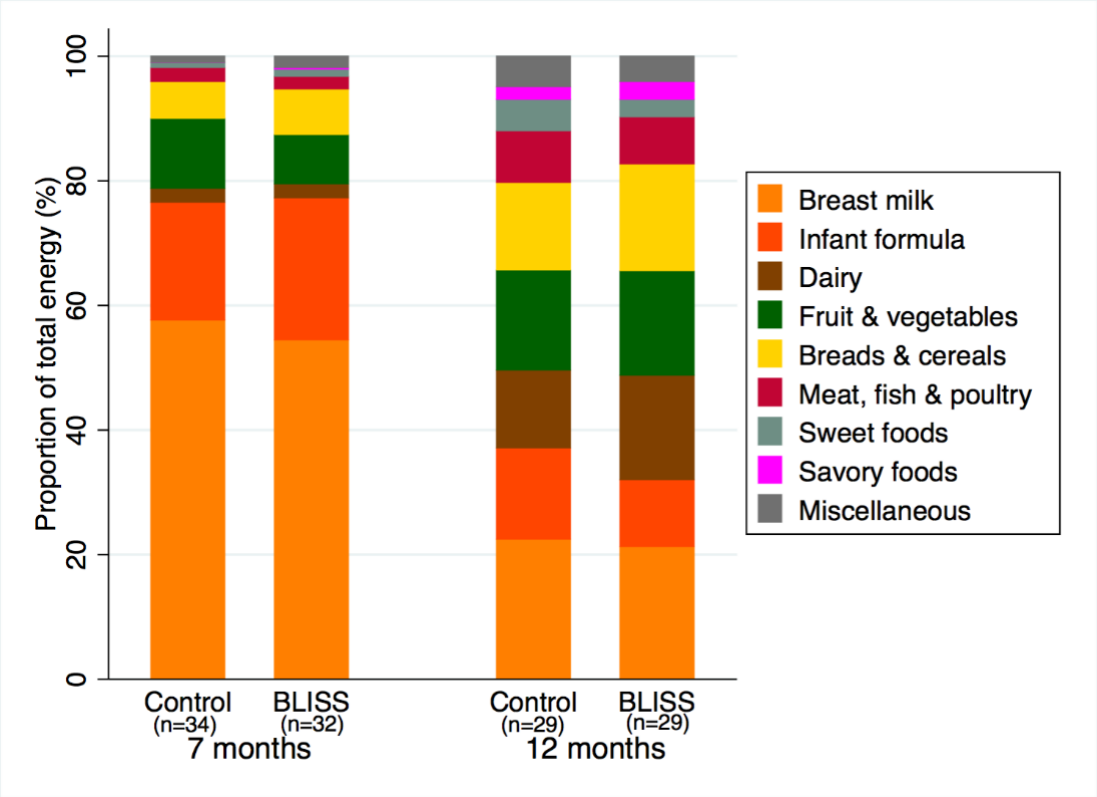


Figure 4.3: Proportion of total energy intake contributed by nine food groups at 7 and 12 months of age, as published in Leong et al. (2018a).

4.3.4 Predictors of alpha diversity

Table 4.7 shows that the only variable associated with alpha-diversity at 12 months of age, from a range of demographic variables, feeding characteristics, and antibiotic use, was parity: i.e. greater alpha diversity was observed in children at higher birth order. The intake of 'infant milk', 'breads and cereals', 'fruit, vegetables, nuts and legumes', and dietary fibre at 7 months of age were all positively and significantly

associated with at least one measure of (higher) alpha diversity at 12 months of age (**Table 4.8**). For example, at 7 months consuming 10 g more of 'breads and cereals' was associated with 9.5 more OTUs and 0.14 higher Shannon Index at 12 months. 'Infant milk' and 'meat' intake were also moderately associated with alpha diversity but this was not statistically significant (number of OTUs: $p = 0.066$ and 0.234, respectively; Shannon Index: $p = 0.089$ and 0.070, respectively).

Table 4.7: Associations between alpha diversity (no. of OTUs and Shannon Index) and variables related to demography, change in feeding mode, and antibiotic use at 12 months (n = 68), as published in Leong et al. (2018a).

	Observed species (Number of OTUs)		**Shannon Index**	
Predictor variables	**Mean difference per unit of predictor variable [a] (95% CI)**	**p-value**	**Mean difference per unit of predictor variable [a] (95% CI)**	**p-value**
Parity (number of children)	23.5 (9.5, 37.5)	**0.001**	0.31 (0.12, 0.49)	**0.002**
Mode of delivery (Caesarean section compared to vaginal)	14.1 (-19.2, 47.4)	0.401	0.17 (-0.27, 0.61)	0.440
Maternal education (has university degree compared to not)	0.3 (-28.9, 29.5)	0.983	-0.03 (-0.41, 0.36)	0.890
Household deprivation (deprivation category)	-13.6 (-34.0, 6.9)	0.189	-0.10 (-0.37, 0.17)	0.473
Consumption of infant formula (had consumed compared to not)	24.2 (-5.2, 53.7)	0.105	0.12 (-0.28, 0.52)	0.547
Age at introduction of infant formula (week)	0.7 (-0.5, 1.9)	0.265	0.01 (-0.002, 0.03)	0.098
Age at introduction of solids (week)	-3.2 (-8.5, 2.1)	0.237	-0.03 (-0.10, 0.04)	0.450
Age at introduction of infant cereal (week)	0.7 (-2.6, 3.9)	0.688	-0.002 (-0.04, 0.04)	0.931
Infant antibiotic usage at time of faecal sample (had consumed compared to not)	15.9 (-29.2, 61.1)	0.483	0.29 (-0.30, 0.88)	0.333

Bold = significant ($p < 0.05$). Abbreviations: OTUs, Operational Taxonomic Units. [a] Unless otherwise indicated. Linear regression with adjustment for group.

Table 4.8: Associations between dietary components at 7 months and alpha diversity at 12 months (no. of OTUs and Shannon Index) (n = 63), as published in Leong et al. (2018a).

		Observed species (Number of OTUs)		Shannon Index	
Dietary components (unit used in regression)	**Intake at 7 months, mean g/d (SD)**	**Mean difference per unit of dietary component [a] (95% CI)**	**p-value**	**Mean difference per unit of dietary component [a] (95% CI)**	**p-value**
Breast milk (100g)	565g (297g)	-2.6 (-7.5, 2.3)	0.300	0.01 (-0.06, 0.07)	0.763
Infant formula (100g)	192g (341g)	3.5 (-0.7, 7.8)	0.100	0.01 (-0.05, 0.06)	0.834
Infant milk (breast or formula) (100g)	757g (135g)	9.8 (-0.7, 20.4)	0.066	0.08 (-0.06, 0.22)	0.234
Breads & cereals (10g)	16.9g (14.9g)	9.5 (0.3, 18.7)	**0.042**	0.14 (0.02, 0.26)	**0.025**
Fruit, vegetables, nuts & legumes (10g)	91.9g (74.7g)	1.8 (-0.2, 3.8)	0.071	0.03 (0.0002, 0.05)	**0.048**
Meat (g)	7.6g (9.6g)	1.3 (-0.2, 2.7)	0.089	0.02 (-0.001, 0.04)	0.070
Dairy [b] (g)	8.5g (10.4g)	0.4 (-0.9, 1.7)	0.565	0.00 (-0.01, 0.02)	0.696
Sweet foods (g)	1.2g (2.8g)	4.0 (-1.2, 9.2)	0.127	0.03 (-0.04, 0.10)	0.337
Savoury foods (g)	0.3g (1.4g)	4.7 (-5.8, 15.3)	0.371	-0.01 (-0.15, 0.13)	0.891
Miscellaneous (10g)	49.1g (88.9g)	0.1 (-1.5, 1.7)	0.933	0.00 (-0.02, 0.02)	0.988
Dietary fibre (g)	3.1g (2.3g)	6.2 (-0.1, 12.5)	0.054	0.10 (0.01, 0.18)	**0.024**
Fibre variety score [c]	11.3 (4.3)	1.03 (-2.8, 4.9)	0.592	0.01 (-0.04, 0.06)	0.672

Bold = significant ($p < 0.05$). Abbreviations: OTUs, Operational Taxonomic Units; g, grams. [a] Linear regression adjusted for group, parity and maternal university education. [b] n = 61. Two participants had very high dairy intake compared to the rest of the sample (98 g and 117 g) and influenced the regression (using the full sample: number of OTUs, mean difference per gram (95% CI) = 0.9 (0.2, 1.6); Shannon Index, mean difference per gram (95% CI) = 0.01 (-0.002, 0.02)). [c] Determined by counting each different fibre containing food (i.e. grain product, vegetable or fruit) consumed over the 3 recording days; n = 56.

Because there were significant associations between diet at 7 months and alpha diversity at 12 months, we hypothesized that diet at 7 months might explain (i.e. mediate) the observed differences in alpha diversity between the BLISS and Control groups. However, as there were no differences between groups for 'infant milk' intake (mean difference, 95% CI: -4, -73 to 65 g/d), 'breads and cereals' intake (mean difference, 95% CI: 1.9, -5.9 to 9.8 g/d) or 'meat' intake (mean difference, 95% CI: -0.9, -6.0 to 4.1 g/d), these foods were not considered to be mediators, leaving only 'fruit, vegetables, nuts and legumes' and dietary fibre intake as potential mediators. **Figure 4.4** explains in detail the mediation model. **Figure 4.5** illustrates the mediating effect that 'fruit, vegetables, nuts and legumes' and dietary fibre intake at 7 months had on the relationship between group and alpha diversity at 12 months. The Control group consumed a mean of 53 g/d more 'fruit, vegetables, nuts and legumes' than the BLISS group and this difference explained 29% and 27% of the association between group and observed species and Shannon Index, respectively. Similarly, the Control group consumed a mean of 1.3 g/d more fibre than the BLISS group, explaining 25% of the relationship between group and both observed species and Shannon Index.

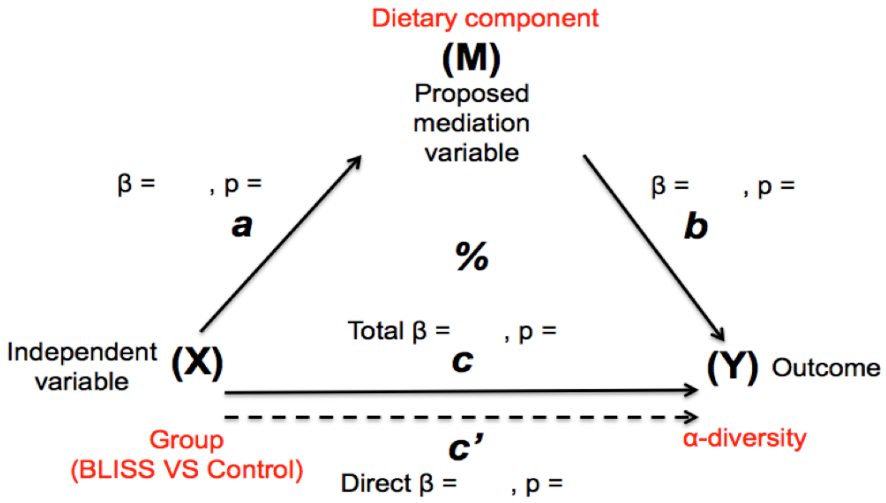


Figure 4.4: Illustration of a single mediation model.

As adapted from (Fairchild & McDaniel, 2017) and (Fairchild & MacKinnon, 2009). Note: X= the independent variable; Y= the dependent variable; and M= the mediating variable. c = ab + c': The mediation model breaks down the total effect of X on Y (i.e. c), into two parts: the indirect effect of X on Y through M, quantified by ab (the product of a and b), and the direct effect of X on Y controlling for M, quantified by c'. %: Represents the proportion of the total effect that is mediated. β is the effect size, and p is the p-value for the effect.

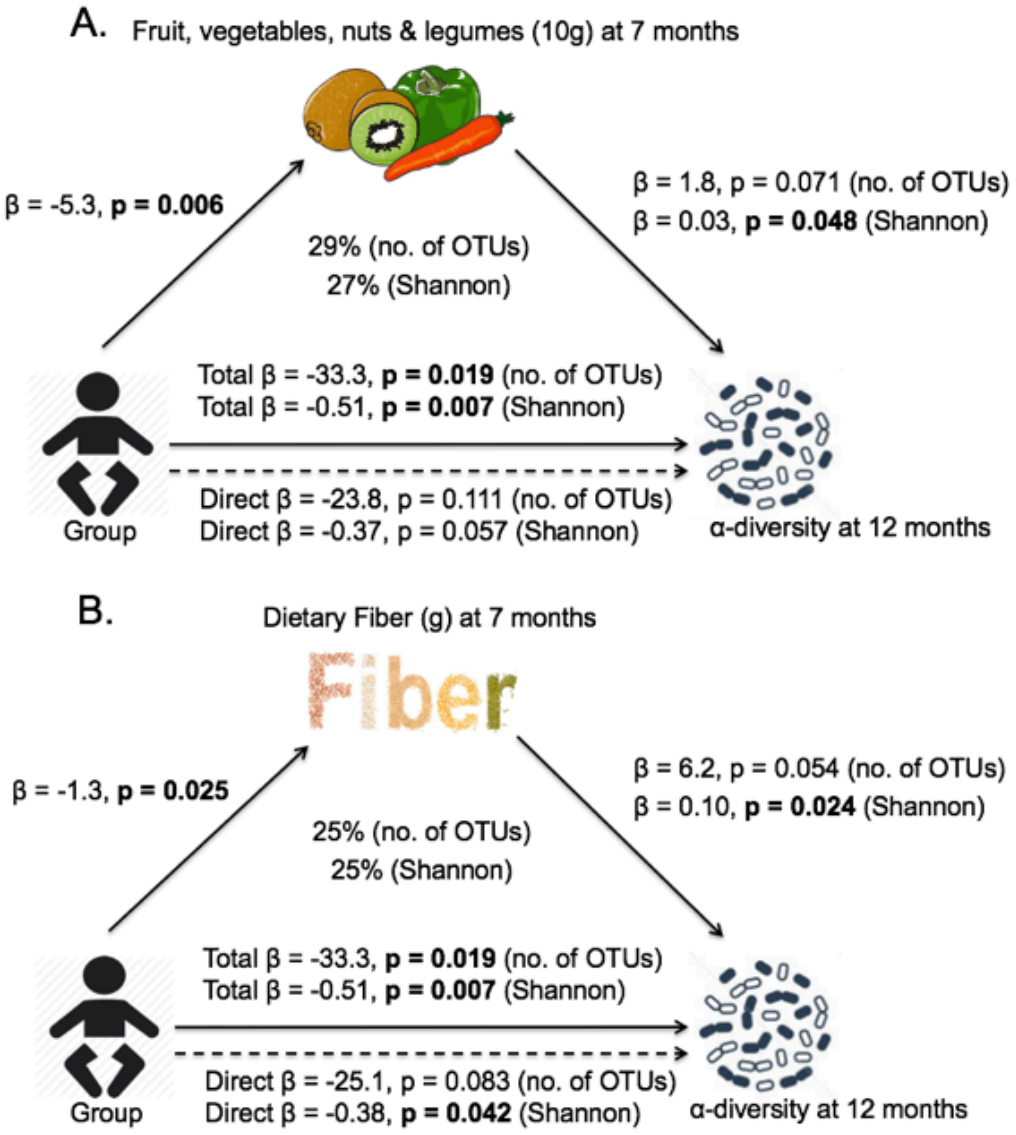


Figure 4.5: The observed relationship between infant group and alpha diversity (mediation model), as published in Leong et al. (2018a).

Bold = significant ($p < 0.05$). (A) Fruit, vegetables, nuts and legumes (for each 10 g). (B) Fibre (for each g) intake. (n = 63). Refer to **Figure 4.4** for explanations of the mediation model.

4.4 Discussion

Microbial composition analysis using several alpha diversity measures revealed that infants following a modified version of BLW, Baby-Led Introduction to SolidS (BLISS), had significantly lower alpha diversity at 12 months of age than Control infants. Using mediation analysis, a novel approach to explore the association between complementary feeding method and alpha diversity, we were able to show that 29%

and 25% of the link between the different methods of complementary feeding and alpha diversity at 12 months could be explained by lower 'fruit, vegetables, nuts and legumes' and 'dietary fibre' intake in the BLISS group at 7 months respectively. There were no differences between the different methods of complementary feeding in the relative abundances of the seven most abundant bacterial families.

Alpha diversity increased from 7 months to 12 months of age in both groups, as expected due to the introduction of sources of food for the gut microbiota other than milk (Krebs et al., 2013; Thompson, Monteagudo-Mera, Cadenas, Lampl, & Azcarate-Peril, 2015). We also noted a decrease in the relative abundance of Bifidobacteriaceae, Enterobacteriaceae and Veillonellaceae, and increases in Lachnospiraceae and Ruminococcaceae from 7 to 12 months. This is largely in agreement with previous studies, which showed similar changes in these bacterial families with increasing infant age (Avershina et al., 2014; Bäckhed et al., 2015; Vallès et al., 2014). The substantial decrease in infant milk intake from 7 to 12 months likely explains the relative decrease in Bifidobacteriaceae which utilise lactose and HMO which are found in breast milk (Sela & Mills, 2010); and the family Veillonellaceae which has species that are able to utilise lactate (Shetty, Marathe, Lanjekar, Ranade, & Shouche, 2013). The relative increase in Lachnospiraceae and Ruminococcaceae at 12 months can be attributed to the introduction of solid foods, as both Lachnospiraceae and Ruminococcaceae have species that are known to play a role in polysaccharide (i.e. 'complex carbohydrate') degradation (Flint et al., 2012a). They form the most abundant bacterial families of the adult faecal microbiota.

To our knowledge, this is the first study to look at the effect of different methods of complementary feeding on the composition of the gut microbiota, the first to investigate the impact of earlier introduction of an adult type diet, and the first to use mediation analysis to determine the pathway of the relationship between a predictor and the gut microbiota. At 12 months of age, infants in the BLISS group had lower bacterial alpha diversity than those in the Control group. The difference was modest, but surprising. The BLISS infants were encouraged to eat the same foods as the rest of the family (i.e. a more adult diet) from the start of complementary feeding and we expected that this would result in greater alpha diversity because a recent study showed that the progression to family foods was strongly associated with increased alpha diversity (observed species and Shannon index) (Laursen et al., 2016). Our results were sustained

even after adjustment for parity, which has been found to be significantly associated with alpha diversity in the current study and elsewhere (Laursen et al., 2015), presumably as a result of exposure to a wider range of bacteria and other environments siblings interact with. Nor can these findings be explained by mode of delivery (Jakobsson et al., 2014) or recent use of antibiotics (Langdon et al., 2016) which have been shown previously to be associated with lower alpha diversity, but were not significant predictors of alpha diversity in the current study.

To investigate how the BLISS intervention was associated with lower alpha diversity, we conducted mediation analyses. Mediation has been described as 'a third-variable effect that explains how or why two variables relate' (Fairchild & McDaniel, 2017). In the current study, mediation analysis allowed us to determine the extent to which the association between the BLISS intervention and alpha diversity was due to differences in specific components of the diet. For mediation to be present, the independent variable (in this case, group) must affect both the outcome (in this case, alpha diversity) and the mediating variable (e.g., food intake); the mediating variable must affect the outcome; and the relationship between the independent variable and the outcome must be reduced in magnitude when the mediating variable is controlled for (Baron & Kenny, 1986; Fairchild & MacKinnon, 2009). In our study, we found that 'fruit, vegetables, nuts and legumes' and 'dietary fibre' intakes were mediators explaining 29% and 25% respectively of the link between the different methods of complementary feeding and alpha diversity (**Figure 4.5**). Even though 'breads and cereals' were associated with alpha diversity, the Control and BLISS groups did not have significantly different intakes of these groups, meaning they were not mediators of the association between group and alpha diversity. The findings suggest that at least a quarter of the impact of the BLISS intervention on alpha diversity can be explained by the lower intake of fruit, vegetables, nuts, legumes and dietary fibre in the BLISS group at 7 months of age. Dietary fibre intake has been shown to be positively associated with alpha diversity measures in this age group (Laursen et al., 2016), presumably because the range of polysaccharides that are indigestible to the host (i.e. fibre) provide growth substrates for a range of gut microbes. This is supported by studies in adults which also demonstrate that dietary fibre is a major driver of diversity (Martinez et al., 2013; Segata, 2015; Tap et al., 2015).

The significantly lower intake of 'fruit, vegetables, nuts and legumes', and dietary fibre, amongst the BLISS infants is an interesting finding. Although there has been an expectation that following a baby-led approach to infant feeding would result in consumption of a wider variety of healthy foods (Rapley & Murkett, 2008), concerns have been expressed that the handheld foods eaten by adults are not necessarily 'healthier' (Cameron et al., 2012b). The infants in the BLISS group were introduced to solids approximately three weeks later than the Control infants so at 7 months of age were earlier in their complementary feeding journey than Controls - perhaps this resulted in them being offered less fruit, vegetables, nuts and legumes at 7 months of age. Moreover, one small pilot study reported that parents who were following BLW with their infant had higher than the recommended intakes of saturated fat, sodium and sugar (Rowan & Harris, 2012), suggesting that if infants were being given the foods that their parents were eating, they may be offered a diet that is less healthy, and therefore probably lower in fruit and vegetables. A more likely explanation for the lower intake of fruit, vegetables, and dietary fibre is that infants who are being traditionally weaned are usually spoon-fed purées in the early months of complementary feeding, and these are commonly based on fruit and vegetables because they can be blended to a smooth consistency and are often sweet and therefore highly palatable. In fact, in the United Kingdom, about two-thirds of all baby foods have been reported to contain fruit, vegetables or both (Garcia, McLean, & Wright, 2016).

Although mediation analysis is increasingly being used in the health literature, this is the first study to our knowledge to use the method to disentangle cause and effect amongst factors associated with microbiological outcomes. In contrast, previous studies in the literature that have looked at mediation have placed the gut microbiota as the mediator ('M' in **Figure 4.4**). Or, in other words, have been interested in whether the gut microbiota helps explains the association between an independent variable and an outcome. For example in one study (Penders et al., 2014), it was mentioned that having older siblings ('X' in **Figure 4.4**) was beneficial to the gut microbiota ('M' in **Figure 4.4**) and that this was linked to a lower risk of atopic dermatitis ('Y' in **Figure 4.4**, i.e. risk of atopic dermatitis was the outcome). In another study (Zhang, Wei, & Chen, 2018), it was reported that intake of fibre ('X' in **Figure 4.4**) was associated with the gut microbiota ('M' in **Figure 4.4**), and that this was linked to body mass index (BMI) ('Y' in **Figure 4.4**, i.e. BMI was the outcome). In contrast, in the current study, we used

food intake as the mediation variable ('M' in **Figure 4.4**) to determine whether it explained the association between the BLISS intervention and the gut microbiota.

Major strengths of our study were the use of a randomised controlled design that ensured that differences between the groups were due to the intervention rather than differences in participant characteristics, and the use of 3-day WDRs to provide detailed high-quality dietary data. Our study also has some limitations. Although we were only able to estimate breast milk intake using total daily volumes from the literature, the number of infants who were breast-fed in each group did not differ at either 7 or 12 months of age. The study was not originally designed to determine the impact of a baby-led approach to complementary feeding on the gut microbiota, and as recruitment started part way through the study, the sample size was small. As the sample size was small, only 9 food groups could be analysed and hence 'legumes and nuts' were placed together with the 'fruit, vegetables, nuts and legumes' food group. It was, however, large enough to identify significant differences in alpha-diversity at 12 months, and we have reported 95% CI so that the reader can determine the range of plausible differences between the groups.

This study shows that infants following a modified version of BLW consume a more adult type diet and have a faecal microbiota with less complex composition at 12 months than infants following traditional spoon-feeding. Lower intakes of 'fruit, vegetables, nuts and legumes' and 'dietary fibre' are partially responsible for this lower alpha diversity. However, the difference in alpha diversity between the different methods of complementary feeding is modest and, at this stage, cannot be related to changes in child development or health. Larger, longer-term studies are required before any conclusions can be made about the possible impact of these differences, or whether infant feeding guidelines should recommend that infants following a baby-led approach to infant feeding consume more fruit, vegetables, nuts and legumes or dietary fibre than is currently the case. This study has however, demonstrated the usefulness of mediation analysis in gut microbiota and diet research.

5 Validation of the EAT5 Food Frequency Questionnaire

This chapter investigates the relative validity and reproducibility of the EAT5 food frequency questionnaire (FFQ) for determining intake of nutrients and food groups of relevance to the gut microbiota. The EAT5 FFQ was used in the POI study (**Chapter 7**). The statistical analysis plan and lessons learnt are included in **Appendix F**.

Chapter highlights:

- Validation of an FFQ for 5-year old children. Few FFQs have been validated for use with young children in the literature.
- Validation of food group intakes of relevance to the gut microbiota. To the best of the Candidate's knowledge, this is the first study to determine in young children, the validity of food groups that are of relevance to the gut microbiota.
- Validation of intake of non-starch polysaccharides (NSP), including both soluble and insoluble NSPs, in young children. Some studies have looked at the validation of total dietary fibre intakes, but none appear to have looked at the ability of an FFQ to measure intake of separate classes of NSPs.

A paper based on this chapter has been published:

Leong, C., Taylor, R. W., Haszard, J. J., Fleming, E., Tannock, G. W., Szymlek-Gay, E. A., Cameron, S. L., Yu, R., Carter, H., Chee, L. K., Kennedy, L., Moore, R., Heath, A.-L. M. (2018). Relative validity and reproducibility of a food frequency questionnaire to assess nutrients and food groups of relevance to the gut microbiota in young children. *Nutrients, 10*(11), 1627. doi: 10.3390/nu10111627

5.1 Introduction

A rapidly expanding literature suggests that the gut microbiota may have beneficial or harmful impacts on health (Clemente, Ursell, Parfrey, & Knight, 2012; Li et al., 2017). Diet plays an important role in modulating gut microbiota, although much of this work has been in adults with little research undertaken in children (Laursen et al., 2017; Lozzo & Sanguinetti, 2018). A dietary component of particular interest is fibre, as it is the main food source for the gut microbiota (Flint, Duncan, & Louis, 2017; Tannock, 2017). Because different classes (soluble, insoluble) or fractions (e.g., arabinoxylan from whole grains, pectin from fruits, and cellulose from vegetables) of fibre appear to impact gut microbiota in different ways (Centanni et al., 2017; Holscher, 2017; Lattimer & Haub, 2010; Leong et al., 2018a), appropriate dietary assessment techniques must be used to improve understanding of how diet influences the microbiota and subsequent health outcomes.

While weighed diet records (WDR) or 24-hour recalls are generally considered gold standard methods of dietary assessment (Gibson, 2005) they entail considerable respondent and researcher burden, and do not directly assess 'usual' intake. FFQs have lower respondent burden, estimate usual intake, and can be used in larger studies examining the long-term effects of diet on the gut microbiota. However, the validity of any new FFQ must be determined in order to ensure that it adequately measures the nutrients of interest in the relevant population (Willett, 1998).

In addition to investigating intake of nutrients, such as fibre, it is also important to be able to determine the intake of foods. This is because most foods are complex combinations of multiple nutrients and food components that cannot be captured by simply measuring nutrient intake, we eat foods rather than nutrients, and dietary guidelines refer to foods rather than nutrients (Willett, 1998).

A handful of validation studies have been undertaken in children evaluating the use of FFQs to assess general dietary intake (Blum et al., 1999; Matos et al., 2012; Moghames et al., 2016; Parrish, Marshall, Krebs, Rewers, & Norris, 2003; Rodríguez et al., 2015; Vereecken, Covents, & Maes, 2010), only fat intake (Dennison, Jenkins, & Rockwell, 2000), or only calcium intake (Huybrechts, De Bacquer, Matthys, De Backer, & De Henauw, 2006; Taylor & Goulding, 1998). However, no studies have validated an FFQ specifically designed to look at both nutrients and food groups of

relevance to the gut microbiota. Hence, the Eating Assessment in Toddlers for 5-year olds (EAT5) validation study provides a good opportunity to validate a tool that would be useful for use in large epidemiological studies investigating gut microbiota in young children.

The objectives addressed in this chapter were to determine:

1. the relative validity and reproducibility of an FFQ for assessing amount and ranking of nutrients of relevance to the gut microbiota; and
2. the relative validity and reproducibility of an FFQ for assessing amount and ranking of food groups of relevance to the gut microbiota.

5.2 Methods

5.2.1 Study design

The study was designed to validate the EAT5 FFQ (**Appendix G**) for measuring intake of nutrients (energy, carbohydrate, dietary fibre, total NSP, soluble NSP, and insoluble NSP) and food groups (i.e. 'higher fibre more healthy cereals', 'higher fibre less healthy cereals', 'lower fibre more healthy cereals', 'lower fibre less healthy cereals', 'nuts and legumes', 'fruits', 'vegetables', 'potatoes and hot chips', 'yoghurt'). Parent and child participants attended two appointments 4 weeks apart. At the first appointment, parents completed the EAT5 FFQ and socio-demographic questionnaire, and anthropometric measurements of the child were obtained. A 3-day WDR was completed over the following 4 weeks. At the second appointment, the EAT5 FFQ was administered again so that reproducibility could be assessed.

5.2.2 Participants

A convenience sample was recruited from Dunedin, Auckland and Wellington (New Zealand) from February 2015 to December 2017. The child had to be healthy and aged 5 to < 6 years during the time of assessment to be eligible for the study. The Human Ethics Committee of the University of Otago, Dunedin, New Zealand, granted ethical approval for the study (reference number H14/154). Written informed consent was obtained from all parents and children.

Parents completed a questionnaire on their child's age, sex, ethnicity and number of siblings. Using the participants' home address, the NZDep2013 Index of Deprivation

was determined (range from 1 to 10, with a value of 1 representing the least deprived 10% of New Zealand households, and a value of 10 representing the 10% most deprived) (Atkinson et al., 2014). The child's height and weight were measured using standard protocols (de Onis, Onyango, Van den Broeck, Chumlea, & Martorell, 2004). Height was measured using a Leicester wall stadiometer (Tanita, Illinois, USA) to the nearest 0.1 cm, with duplicate measures taken (and a third measurement if duplicates were not within 0.7 cm of each other). Weight was measured using digital scales (Seca Alpha model 770; Seca, Hamburg, Germany) to the nearest 0.1 kg, with duplicate measures taken (and a third measurement if duplicates were not within 0.1 kg of each other). Body mass index (BMI) was calculated from the average of height and weight measurements using the formula: weight in kilograms divided by height in metres squared.

5.2.3 Sample size

A sample size of 100 was used based on the recommendation for FFQ validation studies, where a minimum of 50 participants was required, but a sample size of 100 or more participants was preferred (Cade et al., 2002).

5.2.4 Weighed diet record

Parents completed a 3-day WDR on 3 randomly assigned, non-consecutive days (2 week days and 1 weekend day) over 4 weeks. Participants were given detailed verbal and written instructions and a calibrated electronic kitchen scale (Salter Vista, Kent, UK; ± 1 g) at the first visit and then contacted during the collection period so that they could ask further questions. On the second visit, the WDR was collected and checked by 5 Master of Dietetic students and the Candidate. Diet records were analysed with the Kai-culator nutritional software package version 1.16a (Department of Human Nutrition, University of Otago, New Zealand) using the nutrient database FOODfiles 2014 (New Zealand Institute for Plant & Food Research Limited & Ministry of Health, 2014), except for the NSP values where FOODfiles 2010 (New Zealand Institute for Plant & Food Research Limited & Ministry of Health, 2010) was used. Ninety-nine of the 100 3-day WDRs were collected by 5 Master of Dietetics students, as part of their postgraduate studies, with the remaining participant recruited and analysed by the Candidate. Because of the number of data collectors involved, the Candidate developed a protocol for checking and standardizing the individual food item selection in Kai-

culator for all of the 100 3-day WDRs (**Appendix H**). Any change that was made was recorded, and 165 of the 299 diet record days collected were edited by the Candidate.

The WDR also contained a question which asked parents whether the child was unwell on the day of recording, and whether this influenced their child's appetite by increasing or decreasing it. The days of recording where it was indicated that the child was sick and had a decrease in appetite were not used in this thesis because they would be expected to show unrepresentatively low energy reporting that would not be captured by the FFQ. As a result, 11 of 299 (4%) of the diet record days were not used in this thesis.

5.2.5 FFQ development

The EAT5 FFQ had been developed with a particular interest in foods that are relevant to the gut microbiota. It was modified from a validated FFQ used to determine nutrient intakes (Watson et al., 2015) and dietary patterns (Mills et al., 2015) of New Zealand toddlers aged 12 to 24 months. The EAT5 FFQ was designed to be quantitative, interviewer administered, and to capture intake over the past month in 5-year old children. The FFQ had a total of 123 food and beverage items, and included an overall cross-check question for fruit and vegetable frequency (described below).

As the FFQ was meant for older children, infant foods were removed from the original food list. In addition, as the FFQ was designed to look at foods of relevance to the gut microbiota, the food list included questions on a wider range of fruits and vegetables – 14 and 20 items in the EAT5 FFQ compared to 11 and 12 items in the validated FFQ for toddlers.

The FFQ had 10 frequency-response options, ranging from 'not eaten this month' to an open-ended question for multiple times per day. The frequency response options were coded to a per week frequency, with 'less than once a week' coded as 0.5, and the option for 'if more than once a day - how many times a day' multiplied by 7.

Standard units or volume in ml was used to describe portion size. For example, the food item 'kiwifruit' had a unit option of '1 kiwifruit' and the participant was asked 'how many [units] would they eat each time'. In contrast, for the food item 'berries or cherries' the option was how much is eaten in ml, which could be determined using measuring cups or measuring spoons provided by the interviewer.

A cross-check question was used for the fruit and vegetables section: 'how often has your child had fruit (vegetables) in the past month'. A high number of individual questions can lead to overestimates of actual intake. Hence, a cross-check question was considered to be important to help address this problem. A weighting factor was calculated for each individual participant, which was calculated using the following formula: 'number of servings per week from cross-check question' divided by the sum of the 'number of servings per week from individual food items from the related food group (fruit or vegetables)'. For the EAT5 FFQ fruit weighting factor, it would be 'FFQ question 22' divided by 'sum of FFQ question 23-36'. The weighting factor was then used to multiply each individual fruit item frequency to get the fruit and vegetable adjusted (FV-adjusted) frequency value for that individual fruit item. The same was carried out for the vegetable food items (FFQ question 38, 39, 41-58) with the vegetable cross-check question (FFQ question 37). Hot chips were not considered to be vegetables in this thesis (FFQ question 40 and 116) (**Appendix G**).

5.2.6 Entering the FFQ data

Five previous Master of Dietetic students collected 197 of the 199 FFQs for this thesis. The Candidate collected the last 2 FFQs and was responsible for entering all of the EAT5 FFQ (n = 199) data from the hardcopies into an online database (ffq.otago.ac.nz). The Candidate developed a codebook (**Appendix I**) to ensure data consistency, and a protocol for checking the entered data (**Appendix J**). Average values were recorded for those with missing amount values when frequencies but not amounts were reported.

5.2.7 FFQ coding nutrient lines

In order to convert frequency and amounts of foods consumed into nutrients, each FFQ food item entry had to have a 'nutrient line'.

Nutrient lines for the main macro- and micronutrients for the EAT5 FFQ were developed by a Dietitian in consultation with the Candidate and supervisors. Nutrient lines for the EAT5 FFQ were calculated using FOODfiles 2014 (New Zealand Institute for Plant & Food Research Limited & Ministry of Health, 2014) using Kai-culator (Department of Human Nutrition, University of Otago, New Zealand, 2018). For simple FFQ food items such as 'FFQ question 23. Banana', the option 'banana, flesh and seed, raw, ripened, yellow, composite' was chosen as it was the most recent choice with

combined compositions from FOODfiles 2014. For more complex FFQ food items, multiple food options were collapsed into one nutrient line with weightings based on previously reported frequency of consumption and average portion size data of 2 and 10 year olds (Saeedi, Skeaff, Wong, & Skidmore, 2016; Szymlek-Gay, Ferguson, Heath, & Fleming, 2010). For example, 'FFQ question 11. Cornflakes or rice bubbles', the average frequency consumption from the 2 and 10 year olds was 45% for cornflakes, 25% for rice bubbles and 30% for Ricies, and this was used for the weighting. A recipe was created in Kai-culator using 45 g of cornflakes, 25 g of rice bubbles and 30 g of Ricies to determine the nutrient line per 100 g for the 'FFQ question 11. Cornflakes or rice bubbles'. For foods collected by volume, such as cornflakes, a density value was needed so that the nutrient line could be calculated. A flow chart of the steps used to determine these values was developed by the Candidate and used to check the values (**Appendix K)**.

As fibre is an important food source for the gut microbiota (Holscher, 2017; Sawicki et al., 2017), the Candidate developed nutrient lines for fibre fractions for each of the 123 EAT5 FFQ food items. The fibre fractions were: total NSPs, insoluble NSP, soluble NSP, resistant starch, soluble hemicellulose, soluble pectin, insoluble cellulose, insoluble hemicellulose, insoluble pectin and insoluble klason lignin. These fibre fractions are not usually validated in FFQs because so little information is available on fibre fractions for different food items.

However, from the fibre fractions listed above, only the total NSP, insoluble NSP, and soluble NSP could be included in the validation as the Candidate could only be certain of the values for these fibre fractions for the 1010 individual food items from the WDR (the validation study required nutrient values for all the foods reported in the WDRs as well as for the foods in the FFQ). Values for total, insoluble and soluble NSP were obtained from the New Zealand Food Composition Database (FOODfiles 2010, Plant and Food Research) (New Zealand Institute for Plant & Food Research Limited & Ministry of Health, 2010). Similar methods to those used for determining the main nutrients were used to determine the fibre fraction nutrient lines for the 123 FFQ food items. **Appendix L** shows an example of some of the decisions made. The analytical method used to determine total dietary fibre, total, insoluble and soluble NSP in FOODfiles is the enzymatic-gravimetric method, Association of Analytical

Communities (AOAC) 991.43 (Horwitz & Latimer, 2006; New Zealand Institute for Plant & Food Research Limited & Ministry of Health, 2010).

a. FFQ coding of other fibre fractions - not used in the FFQ validation (i.e. resistant starch, soluble hemicellulose, soluble pectin, insoluble cellulose, insoluble hemicellulose, insoluble pectin and insoluble klason lignin)

The other fibre fraction values that were not available in FOODfiles, and so could not be generated for the 1010 individual foods in the WDRs were determined only for the FFQ food items. The literature provides values for a relatively limited list of foods; hence, values were developed for the 123 EAT5 FFQ food items, but could not be developed for the 1010 individual food items from the WDRs. For resistant starch values, most of the FFQ food item values were obtained from the Australian Resistant Starch Report (Landon et al., 2012) which used the AOAC 2002.02 and AACC32-40 methods. For the 7 food items that could not be found in the report, values were obtained from a database of resistant starch amounts in foods in the United States (Murphy et al., 2008). These values were a mean of many values from the literature that used a variety of analytical methods (Murphy et al., 2008).

Values for soluble hemicellulose and pectin; and insoluble cellulose, hemicellulose, pectin and klason lignin were calculated from the database by Marlett and Cheung for 228 commonly consumed foods (Marlett & Cheung, 1997). The database uses total fibre values determined using the Uppsala method (Theander, Åman, Westerlund, & Graham, 1990). The Uppsala method has been shown to be comparable to the AOAC enzymatic-gravimetric method for vegetables (Marlett & Vollendorf, 1993), fruits (Marlett & Vollendorf, 1994a) and cereals and grains (Marlett & Vollendorf, 1994b). Similar methods as used for determining the main nutrients were used to determine the fibre fraction nutrient lines for the 123 FFQ food items.

For future research use, an example of a flow chart for the steps taken to develop the final nutrient lines for these other fibre fractions for the EAT5 FFQ food items can be found in **Appendix M**.

5.2.8 FFQ food groups

Twelve food groups were defined based on food groups that were considered to be of relevance to the gut microbiota (as described below), and the number of

consumers (i.e. at least 8 consumers were required in each food group so that there was sufficient power to perform the food group analyses (Floyd & Widaman, 1995)). The FFQ food items were allocated to the food groups based on nutrient profile and similarity of use (**Figure 5.1**). **Table 5.1** shows the starch and fibre fraction content of each food group, calculated as the mean grams of the nutrient/100 g of the foods in the food group. The following decisions were made when creating the food groups:

- Carbohydrate foods were separated into 4 groups. 'Lower fibre' means fibre less than 3.4 g per 100 g. The cut-off was based on the median value for the 1010 individual foods reported in the WDRs. 'Less healthy' means foods that are not staple carbohydrates, foods that are eaten as snacks, or foods that are high in saturated fat or high in sugar (>15 g /100 g) (Cooper, Martin, & Keim, 2015; Keim & Martin, 2014; So et al., 2018).
- Nuts, seeds and legumes are high in fibre, but have a different nutrient profile to grains and cereals, or fruit and vegetables. Moreover, a recent study on nuts has shown that they have prebiotic compounds that may affect the gut microbiota (Fernando et al., 2010; Lamuel-Raventos & Onge, 2017).
- Potatoes and hot chips were separated from vegetables as they contain more starch than vegetables, and are the food group with the highest amount of resistant starch per 100 g (Lyte et al., 2016).
- Yoghurt was separated from dairy as it is likely to contain probiotics (Singh et al., 2017).

The same 12 food groups were applied to foods reported in the 3-day WDRs. Each of these 1010 individual foods was assigned to one of the 12 food groups, except for water, which was excluded from the food groups. Similar to the food group allocation method described in section 3.5.3, an Excel spreadsheet was created with the energy (kJ) per 100g for each of the 1010 individual food items (similar to the spreadsheet in **Appendix C**). For each food item the Candidate determined the contribution of that food item to the energy from each of the 12 food groups (per 100g). The Candidate also carried out technical quality control checks on the exported data from Kai-culator.

Food group	EAT5 food frequency questionnaire food item
Higher fibre (≥3.4g/100g) more healthy cereals	Wholegrain bread or bun, Wholemeal bread or bun, White buns, Pizza (not takeaway), Crackers (wheat, rice or corn-based), Rice cakes or rice wheels, Cruskits or crispbreads, Weet-bix
Higher fibre (≥3.4g/100g) less healthy cereals	Fruity-bix or similar, Muesli and light muesli, Other breakfast cereal, Fruit bread, Currant buns, Muesli or nut or cereal or puffed rice bars, Crisps, Corn chips, Corn snacks (e.g. Cheezels)
Lower fibre more healthy cereals	White bread, White rice, Instant noodles, Canned spaghetti, Other pasta (not including sauce), Brown rice, Porridge, Cornflakes or rice bubbles
Lower fibre less healthy cereals	Cocopops, Honey puffs or puffed wheat cereal, Nutrigrain, Milo cereal or similar, Biscuits - chocolate coated, Biscuits – other, Cakes or slices, Muffins or scones, Croissant, Sweet buns, Iced buns, Pastries, Puddings not yet described
Nuts and legumes	Hummus (chickpea dip), Baked beans, Canned or home cooked beans, Chickpeas or lentils, Peanut butter, Nuts (any sort but not peanut butter)
Fruits	Apples (fresh or canned), Pears, Banana, Raisins or sultanas, Dried apricots or prunes, Apricots, Plums, Peaches, Oranges, Mandarins, Kiwifruit, Green grapes, Black or red grapes, Berries or cherries, Avocado, Rhubarb, Other fruit
Vegetables	Yams, Carrot, Pumpkin, Green peas, Green beans, Sweet corn, Broccoli, Cauliflower, Capsicum (peppers), Red cabbage, Green cabbage, Spinach or silverbeet, Lettuce or salad leaves, Cucumber, Raw tomato, Cooked tomato (pasta sauce, canned tomatoes, tomato sauce), Leeks, Other vegetables
Potatoes and hot chips	Potato salad or other potato eaten cold, Potato or kumara (boiled, baked, microwaved, mashed) eaten warm or hot, Hot chips, Potato shapes, Roast potato or kumara cooked at home eaten warm or hot, Chips from a takeaway shop or fast food restaurant
Dairy	Cow's milk as a drink, Cow's milk on cereal or other food, Low-fat cow's milk as a drink, Low-fat cow's milk on cereal or other food, Cheese (including in recipes), White sauce or cheese sauce, Cream or sour cream, Ice cream, Custard and other milk puddings, Flavoured milk (including Milo, Quick, Drinking chocolate, Up-and-Go)
Yoghurt	Yoghurt or dairy food
Meat, fish, egg	Eggs, Mince & patties (from beef or lamb), Steak, Chops or roast (beef or lamb), Fish (e.g. canned, pan-fried), Chicken (e.g. roast, stir-fry, BBQ), Pork and other meat, Sausages, saveloys, Ham, Bacon, Luncheon, Salami
Miscellaneous	100% orange juice (freshly squeezed or similar), Other fruit juice ('Fresh up', 'Just Juice'), Soy milk as a drink, Soy milk on cereal or other food, Other milk (goat, rice) as a drink, Other milk (goat, rice) on cereal or other food, Fruit drinks, Ribena, Cordial, Sachets, Regular fizzy drinks (lemonade, coke), Diet fizzy drinks (lemonade, coke), Tea (not herbal), Coffee, Jam or honey, Marmite or vegemite, Nutella, Butter (not in baking), Margarine (not in baking), Chocolate, Lollies, Fruit leather, Fruit strings fruit roll-ups, Meat pies, Sausage rolls, Fish fingers or shapes, Battered or crumbed fish, Chicken nuggets or shapes, Hotdog or fish or sausage from a takeaway shop, Burgers from a takeaway shop or fast food restaurant, Other item from a takeaway shop or fast food restaurant, Ready to eat pizza (takeaway shop or supermarket), KFC or other fried chicken, Subway sandwich, Kebabs or wraps (bought), Sushi (bought), Chinese, Thai or Indian meal or similar (bought)

Figure 5.1: Food groups developed based on their relevance to the gut microbiota, as published in Leong et al. (2018b).

Table 5.1: Starch and fibre fraction composition of food groups per 100 g.

	Nutrients in g/100 g											
					Soluble			Insoluble				
Food group	**Starch**	**Resistant starch**	**Fibre**	**Total NSP**	**NSP**	**Hemicellulose**	**Pectin**	**NSP**	**Cellulose**	**Hemicellulose**	**Pectin**	**Klason lignin**
Higher fibre more healthy cereals	44.3	1.0	7.0	5.4	2.4	1.0	0.0	3.0	1.8	2.8	0.0	0.7
Lower fibre more healthy cereals	28.0	1.0	1.6	1.4	0.9	0.3	0.0	0.5	0.4	0.4	0.0	0.3
Higher fibre less healthy cereals	39.0	0.7	7.1	4.5	2.4	0.7	0.1	1.8	1.7	2.9	0.2	0.7
Lower fibre less healthy cereals	25.3	0.5	1.9	1.3	0.7	0.4	0.0	0.6	0.5	0.5	0.0	0.3
Nuts and legumes	8.5	0.8	6.1	4.6	3.3	0.6	0.2	1.3	1.9	1.6	0.7	0.5
Fruits	0.6	0.1	2.7	2.2	1.2	0.2	0.3	1.0	0.6	0.5	0.3	0.5
Vegetables	2.0	0.2	2.4	2.1	1.1	0.1	0.2	0.9	0.8	0.5	0.4	0.2
Potatoes	19.3	1.3	2.5	2.3	1.1	0.2	0.2	1.3	0.8	0.5	0.2	0.0
Dairy	0.5	0.0	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
Yoghurt	0.9	0.0	0.2	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
Meat, fish, egg	0.8	0.0	0.3	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0
Miscellaneous	8.6	0.0	1.2	1.2	0.5	0.1	0.0	0.7	0.1	0.1	0.0	0.1

5.2.9 Statistical analysis

Data were analysed using Stata software (StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP).

The FFQ frequencies and amounts were collected on a per week basis so were divided by 7 to get the average daily intake for assessing energy, nutrient and food group intakes. This was compared with the average daily intake from the WDR to determine the relative validity of the FFQ. The FFQ used for this comparison was a randomly selected first or second FFQ for each participant (except for the participant who provided only a first FFQ). The WDR data were adjusted for intra-individual variation using the Multiple Source Method (MSM) program (Harttig et al., 2011) in order to provide a better estimate of 'usual intake'.

Histograms were plotted for each variable and used to visually assess the normality of their distribution. The majority of the distributions were right-skewed so geometric means and 95% confidence intervals (CI) were used to describe the intakes. However, the majority of the distributions of the paired differences (i.e. between the two administrations of the FFQ) were normally distributed, and hence the arithmetic mean difference was reported using a paired t-test. Spearman's correlation coefficients were calculated comparing the FFQ with the WDR. Correlations of 0.30 - 0.49 were considered 'acceptable', 0.50 - 0.70 'good' (Willett, 1998), and > 0.70 were considered 'very good'.

Cross-classification of WDR and FFQ quartiles was also carried out. The percentage of participants correctly classified was defined as the FFQ categorizing the diet into the same quartile as the WDR, while gross misclassification was defined as the FFQ categorizing the diet into the highest quartile when the WDR was categorized into the lowest quartile, and vice versa. The 'actual values for surrogate categories' approach determines the extent to which intakes measured using a new method (in this case the EAT5 FFQ) reflect intakes measured using a reference method (WDR). Actual values for surrogate categories (Willett, 1998) were calculated as follows: participants were assigned to quartiles according to intake estimated by the EAT5 FFQ, then the mean intake in each quartile was calculated using the intake determined by the WDR. Regression analyses were used to see whether there was a trend in the step-wise increases across the quartiles, and a difference between quartile 1 and quartile 4.

Bland-Altman analyses and plots (Bland & Altman, 1999) were used to assess the agreement between the FFQ and WDR at the individual level.

Intra-class correlation coefficients were calculated comparing the first and second administration of the EAT5 FFQ to assess reproducibility, with correlations of 0.30-0.49 considered 'acceptable', 0.50-0.70 'good' (Willett, 1998), and >0.70 considered 'very good'.

Data were reported for both the crude EAT5 FFQ values and also for FV-adjusted EAT5 FFQ values, which used the weighting factor for the fruit and vegetable food items (as discussed in section 5.2.5).

5.3 Results

5.3.1 Study population

One hundred participants were recruited, of whom 99 parent-child pairs completed the two FFQs and the 3-day WDR. One parent-child pair completed only the first FFQ and the 3-day WDR, meaning that 100 participants were included in the validity analysis and 99 participants in the reproducibility analysis. The 100 young children (44% male) had a mean (range) age of 5.5 (4.9 - 6.0) years and BMI of 16.0 (13.7 - 19.7) kg/m^2 (**Table 5.2**). The participants were mainly of New Zealand European ethnicity (80%), with 13% Māori and 5% Asian. Twenty percent of the participants had no siblings. According to the NZDep2013 Index of Deprivation (Atkinson et al., 2014), 19% of the participants were from households in the three most deprived deciles (compared to the expected 30%).

Table 5.2: Characteristics of the EAT5 study participants (n = 100).

Variable [a]	
Male, n (%)	44 (44)
Age, (year)	5.5 (0.3)
Weight, (kg)	20.3 (2.3)
Height, (cm)	112.7 (4.7)
BMI, (kg/m^2)	16.0 (1.1)
Parity, n (%)	
1	19.6 (19.6)
>1	80.4 (80.4)
Ethnicity, n (%)	
NZ European	80 (80)
Māori/Pasifika	13 (13)
Asian	5 (5)
Other	2 (2)
Household deprivation decile [b], n (%)	
1-3 (low)	46 (46)
4-7	34 (34)
8-10 (high)	19 (19)

Abbreviations: EAT, Eating Assessment in Toddlers; BMI, body mass index (calculated as weight in kilograms divided by height in meters squared). [a] Data presented as mean (SD) unless otherwise indicated. Data were missing for 3 participants for parity and 1 for household deprivation. [b] Determined using the New Zealand Index of Deprivation 2013 (Atkinson et al., 2014). The Index combines 9 variables from the 2013 New Zealand National Census to provide a deprivation score for each mesh block (a geographical unit defined by Statistics New Zealand that contains approximately 81 people). The score reflects the extent of material and social deprivation and is used to construct deciles from 1 (low deprivation) to 10 (high deprivation).

5.3.2 Nutrient intake

In general, the FV-adjusted EAT5 FFQ data were closer to the WDR data than the crude unadjusted EAT5 FFQ data, so these data are reported here. For easier reference, from now on and in the Tables, the FV-adjusted EAT5 FFQ will be referred to as 'FV-FFQ' and the crude unadjusted EAT5 FFQ will be referred to as 'Crude FFQ', and 'FFQ' on its own refers to both 'FV-FFQ' and 'Crude FFQ'. The FFQ is a randomly selected FFQ from either the first or second administration.

Table 5.3 shows the energy and selected nutrient estimates from the FFQ compared to the WDR. The average energy intake of 5845 kJ reported for these diet records was much lower than that for the New Zealand national survey of 5-6 year olds, where males had a median intake of 7573 kJ and females 6703 kJ (Ministry of Health, 2003). There were no significant differences in mean carbohydrate, fibre and total NSP intakes measured by the FV-FFQ and the WDR. Estimates of intakes of macro- and micronutrients of less relevance to the gut microbiota can be found in **Table 5.4**.

Table 5.3: Mean daily intake, mean difference and limits of agreement for selected nutrients of relevance to the gut microbiota according to WDR and FFQ in 5-year old children (n = 100).

Nutrient	WDR		Crude FFQ [a]		FV-FFQ [a]		FV-FFQ [a] vs WDR			
	Mean [b]	(95% CI)	Mean [b]	(95% CI)	Mean [b]	(95% CI)	Mean diff	(95% CI)	p-value [c]	LOA [d]
Energy (kJ)	5845	(5613, 6086)	6866	(6488, 7267)	6476	(6107, 6866)	792	(398, 1187)	<0.001	-3183-4768
Carbohydrate (g)	181	(173, 190)	201	(191, 213)	186	(175, 197)	8.1	(-3.3, 19.4)	**0.163**	-107-123
Dietary fibre (g)	18	(17, 19)	23	(22, 25)	19	(18, 20)	1.2	(-0.2, 2.6)	**0.090**	-13-15
Total NSP (g)	15	(14, 16)	19	(18, 20)	16	(15, 17)	0.4	(-0.7, 1.5)	**0.488**	-11-12
Soluble NSP (g)	6.3	(5.9, 6.6)	9.9	(9.2, 10.7)	8.0	(7.4, 8.6)	2.0	(1.4, 2.7)	<0.001	-4.6-8.6
Insoluble NSP (g)	8.9	(8.3, 9.6)	9.1	(8.5, 9.7)	7.5	(7.0, 8.0)	-1.6	(-2.3, -1.0)	<0.001	-8.0-4.7

Bold = not statistically significantly different at $p < 0.05$. Abbreviations: CI, confidence interval; diff, difference; LOA, limits of agreement; NSP, non-starch polysaccharides. [a] FFQ1 or FFQ2 was randomly chosen for each participant. [b] Geometric mean. [c] Paired t-test. [d] Bland-Altman limits of agreement.

Table 5.4: Mean daily intake, mean difference and limits of agreement for other nutrients according to WDR and FFQ in 5-year old children (n = 100).

Nutrient	WDR		Crude FFQ [a]		FV-FFQ [a]		FV-FFQ [a] vs WDR			
	Mean [b]	(95% CI)	Mean [b]	(95% CI)	Mean [b]	(95% CI)	Mean diff	(95% CI)	p-value [c]	LOA [d]
Protein (g)	51	(49, 53)	65	(61, 70)	63	(59, 67)	14	(9.6, 18)	<0.001	-28-55
Total fat (g)	50	(47, 53)	62	(58, 66)	59	(56, 64)	11	(6.7, 16)	<0.001	-35-58
Sugar (g)	77	(72, 82)	94	(87, 100)	82	(76, 89)	6.8	(0.4, 13.3)	0.039	-58-72
Sodium (mg)	1560	(1453, 1675)	1901	(1790, 2020)	1793	(1686, 1908)	233	(95, 371)	0.001	-1155-1621
Calcium (mg)	593	(548, 640)	730	(669, 795)	686	(628, 751)	123	(58, 187)	<0.001	-532-777
Iron (mg)	8.8	(8.2, 9.5)	10.6	(9.9, 11.3)	9.8	(9.1, 10.5)	0.9	(0.1, 1.6)	0.021	-6.4-8.2
Vitamin C (mg)	63	(56, 70)	107	(95, 119)	74	(65, 84)	16	(6.1, 25)	0.002	-81-113

Bold = not statistically significant at $p < 0.05$. Abbreviations: CI, confidence interval; diff, difference; LOA, limits of agreement; NSP, non-starch polysaccharides. [a] FFQ1 or FFQ2 was randomly chosen for each participant. [b] Geometric mean. [c] Paired t-test. [d] Bland-Altman limits of agreement.

The mean correlation between nutrients measured by the FV-FFQ and WDR was 0.34 ('acceptable'), with a range from 0.24 for soluble NSP to 0.38 for total and insoluble NSP (**Table 5.5**). The correlations were slightly higher for MSM adjusted values with a mean of 0.35 ('acceptable'). **Table 5.5** also reports the correlations used to assess reproducibility of the first and second administration of the FV-FFQ. The mean correlation was 0.83 ('very good'), with a range from 0.80 to 0.88. The reproducibility correlations for the other nutrients can be found in **Table 5.6**.

Table 5.5: Selected nutrients correlations between the FFQ and WDR (n = 100), and reproducibility (n = 99) in 5-year old children.

Nutrient	Relative validity [a]				Reproducibility [b]	
	Crude FFQ [c] vs WDR	FV-FFQ [c] vs WDR	MSM adjusted FV-FFQ [d] vs WDR	Previous studies [e]	FV-FFQ1 vs FV-FFQ2	Previous studies [f]
Energy (kJ)	0.32	0.32	0.32	0.19-0.66	0.88	0.29-0.73
Carbohydrate (g)	0.36	0.37	0.37	0.14-0.66	0.87	0.26-0.67
Dietary fibre (g)	0.33	0.36	0.38	0.02-0.60	0.80	0.26-0.78
Total NSP (g)	0.32	0.38	0.39	NR	0.80	NR
Soluble NSP (g)	0.18	0.24	0.25	NR	0.80	NR
Insoluble NSP (g)	0.34	0.38	0.39	NR	0.80	NR

Abbreviations: NSP, non-starch polysaccharides; NR, not reported. [a] Spearman's correlation coefficients. [b] Intra-class correlation coefficients. [c] FFQ1 or FFQ2 was randomly chosen for each participant. [d] Using Multiple Source Method (MSM) to adjust for the intra-individual variation occurring between the 3 days of diet records. [e] Inclusion of 12 studies with correlations for nutrients: three Spearman's correlations (Marshall, Eichenberger Gilmore, Broffitt, Levy, & Stumbo, 2003; Moghames et al., 2016; Vereecken et al., 2010), nine Pearson's correlations (Blum et al., 1999; Delisle Nyström et al., 2017; Dennison et al., 2000; Fumagalli, Pontes Monteiro, Sartorelli, Vieira, & de Lourdes Pires Bianchi, 2008; Kobayashi et al., 2011; Matos et al., 2012; Rodriguez et al., 2017; Stein, Shea, Basch, Contento, & Zyberf, 1992; Vioque et al., 2016). [f] Inclusion of 5 studies with correlations for reproducibility for nutrients: 1 Intra-class correlation coefficients (Moghames et al., 2016), 4 Pearson's correlations (Dennison et al., 2000; Kobayashi et al., 2011; Rodriguez et al., 2017; Vioque et al., 2016).

Table 5.6: Other nutrient correlations between the FFQ and WDR (n = 100), and reproducibility (n = 99) in 5-year old children.

Nutrient	Relative validity [a]		Reproducibility [b]
	Crude FFQ vs WDR [c]	FV-FFQ vs WDR [c]	FV-FFQ1 vs FV-FFQ2
Protein (g)	0.43	0.41	0.87
Total fat (g)	0.25	0.26	0.85
Sugar (g)	0.36	0.42	0.88
Sodium (mg)	0.24	0.23	0.87
Calcium (mg)	0.49	0.50	0.90
Iron (mg)	0.46	0.45	0.89
Vitamin C (mg)	0.51	0.49	0.81

[a] Spearman's correlation coefficients. [b] Intra-class correlation coefficients. [c] FFQ1 or FFQ2 was randomly chosen for each participant.

All nutrients and energy had a percentage correctly classified into quartiles by the FV-FFQ and WDR that was greater than chance (25%), ranging from 28% (dietary fibre and insoluble NSP) to 36% (carbohydrate) (**Table 5.7**). The mean percentage grossly misclassified was 5.7%, and correctly classified to extreme quartiles was 19.1% (12.5% would be expected by chance alone). The percentage cross-classifications for the other nutrients can be found in **Table 5.8**.

Trends for the actual values for surrogate categories show the expected increase across the FV-FFQ quartiles for energy and all nutrients (all $p \leq 0.005$) (**Table 5.9**). The FV-FFQ clearly differentiated between the first and fourth quartile for energy and all nutrients (all differences between the first and fourth quartile $p \leq 0.011$).

Table 5.7: Selected nutrient cross-classifications between the FFQ and WDR in 5-year old children (n = 100).

Nutrient	Cross-classification (FFQ [a] vs WDR)							
	% Correctly classified [b]		% Correct & adjacent [c]		% Grossly misclassified [d]		% Correct extremes [e]	
Chance	*25%*		*62.5%*		*12.5%*		*12.5%*	
	Crude	FV	Crude	FV	Crude	FV	Crude	FV
Energy (kJ)	32	34	79	79	9	9	18	19
Carbohydrate (g)	38	36	77	76	5	6	25	23
Dietary fibre (g)	34	28	79	76	5	4	20	18
Total NSP (g)	28	29	75	77	9	6	16	18
Soluble NSP (g)	28	32	69	69	9	5	19	19
Insoluble NSP (g)	38	28	75	76	7	4	20	18

Abbreviations: NSP, non-starch polysaccharides. [a] FFQ1 or FFQ2 was randomly chosen for each participant. [b] % Correctly classified = percentage of children with WDRs and FFQ intakes in the same quartile. [c] % Correct and adjacent = percentage of children with WDRs and FFQ intakes in the same and 1 adjacent quartile. [d] % Grossly misclassified = percentage of children with WDRs intakes in the highest quartile and FFQ intakes in the lowest quartile and vice versa. [e] % Correctly classified to extreme quartiles = percentage of children with WDRs and FFQ intakes correctly classified to the lowest and highest quartiles.

Table 5.8: Other nutrient cross-classifications between the FFQ and WDR in 5-year old children (n = 100).

Nutrient	Cross-classification (FV-FFQ [a] vs WDR)			
	% Correctly classified [b]	% Correct & adjacent [c]	% Grossly misclassified [d]	% Correct extremes [e]
Chance	*25%*	*62.5%*	*12.5%*	*12.5%*
Protein (g)	42	78	8	24
Total fat (g)	29	67	6	20
Sugar (g)	40	78	4	24
Sodium (mg)	32	71	9	17
Calcium (mg)	43	84	3	27
Iron (mg)	39	78	5	24
Vitamin C (mg)	31	78	1	22

[a] FFQ1 or FFQ2 was randomly chosen for each participant. [b] % Correctly classified = percentage of children with WDRs and FFQ intakes in the same quartile. [c] % Correct and adjacent = percentage of children with WDRs and FFQ intakes in the same and 1 adjacent quartile. [d] % Grossly misclassified = percentage of children with WDRs intakes in the highest quartile and FFQ intakes in the lowest quartile and vice versa. [e] % Correctly classified to extreme quartiles = percentage of children with WDRs and FFQ intakes correctly classified to the lowest and highest quartiles.

Table 5.9: Ability of the EAT5 FFQ to differentiate between quartiles of WDR intake, determined using actual values for surrogate categories (n = 100), as published in Leong et al. (2018b).

Nutrient	*Quartiles defined by* [a]	Mean WDR intake				p-value for trend [b]	p-value for Q1 vs Q4 [c]
		Q1	***Q2***	***Q3***	***Q4***		
Energy (kJ)	FV-FFQ	5414	5858	6321	6262	**0.004**	**0.001**
	WDR	4527	5505	6346	7477		
Carbohydrate (g)	FV-FFQ	164	186	193	199	**0.002**	**0.003**
	WDR	132	175	198	238		
Dietary fibre (g)	FV-FFQ	16	18	19	21	**0.005**	**0.007**
	WDR	12	16	20	27		
Total NSP (g)	FV-FFQ	14	15	17	18	**0.001**	**0.002**
	WDR	10	14	17	23		
Soluble NSP (g)	FV-FFQ	5.7	6.6	6.6	7.2	**0.001**	**0.008**
	WDR	4.4	5.7	6.8	9.3		
Insoluble NSP (g)	FV-FFQ	8.1	9.1	9.5	11	**0.001**	**0.001**
	WDR	5.7	8.2	10	14		

Bold = significant ($p < 0.05$). Abbreviations: NSP, non-starch polysaccharides. [a] FFQ1 or FFQ2 was randomly chosen for each participant. [b] Significant difference in the trend across the quartiles (regression). [c] Significant differences between Q1 vs Q4 (regression).

However, the limits of agreement between the FFQ and the WDR were wide for energy and all nutrients (**Table 5.3** and **Table 5.4**). The Bland-Altman plots show little bias in the FV-FFQ with a good scatter in the plots for energy and nutrients (**Figure 5.2**), although possible overestimation of energy and soluble NSP at higher intakes was indicated.

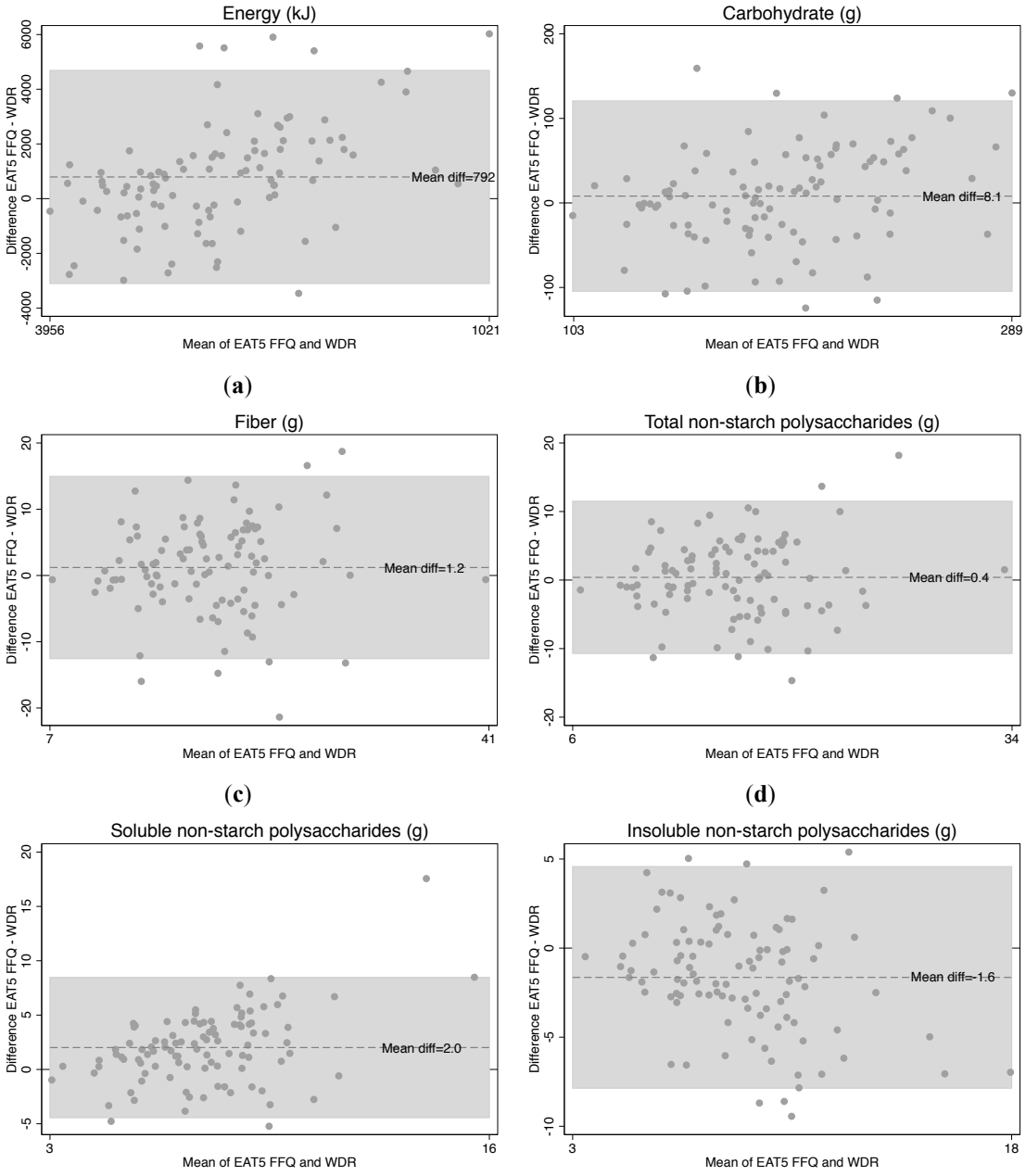


Figure 5.2: Bland-Altman plots for nutrient intakes from a randomly chosen FV-FFQ1 or FV-FFQ2 and the WDR, as published in Leong et al. (2018b).

(a) Energy in kJ; (b) Carbohydrate in grams; (c) Fibre in grams; (d) Total non-starch polysaccharides in grams; (e) Soluble non-starch polysaccharides in grams; and (f) Insoluble non-starch polysaccharides in grams.

5.3.3 Food group intake

For the food groups, only the FV-adjusted EAT5 FFQ data is shown and discussed.

Table 5.10 shows the mean energy contribution from each food group for the FV-FFQ compared to the WDR. For 6 of the 12 food groups, the mean FV-FFQ intake was not significantly different from the WDR. For the other 6 foods groups, the FV-FFQ gave significantly higher estimates than the WDR. Similarly, estimates of the mean amount eaten in grams from each food group can be found in **Table 5.11**.

The mean correlation for food group intakes between the FV-FFQ and WDR was 0.41 ('acceptable'), with a range from 0.28 for 'vegetables' and 'miscellaneous' to 0.56 ('good') for 'meat, fish, egg' (**Table 5.12**). **Table 5.12** also reports the correlations used to assess reproducibility of the estimates of food group intake between the first and second administration of the FV-FFQ. The mean correlation was 0.80 ('very good'), with a range from 0.57 ('potatoes': 'good') to 0.91 ('lower fibre more healthy cereals': 'very good'). Similarly, the relative validity and reproducibility correlations for the amount eaten in grams from each food group can be found in **Table 5.13**.

All food groups had a percentage correctly classified into quartiles by the FV-FFQ and WDR that was greater than chance (25%), with a range from 28% ('vegetables') to 51% ('meat, fish, egg') (**Table 5.14**). The mean percentage grossly misclassified was 5.1% and correctly classified to extreme quartiles was 22.8% (12.5% would be expected by chance alone). Similarly, the percentage cross-classifications for the amount eaten in grams from each food group can be found in **Table 5.15**.

Table 5.10: Mean daily intake, mean difference and limits of agreement for food groups (energy contribution) of relevance to the gut microbiota according to WDR and FV-FFQ in 5-year old children (n = 100), as published in Leong et al. (2018b).

Food group	WDR		FV-FFQ [a]		FV-FFQ [a] vs WDR			
	Mean [b]	(95% CI)	Mean [b]	(95% CI)	Mean diff	(95% CI)	p-value [c]	LOA [d]
Higher fibre more healthy cereals (kJ)	858	(740, 996)	805	(689, 941)	-22	(-161, 118)	**0.758**	-1428-1385
Lower fibre more healthy cereals (kJ)	498	(430, 577)	422	(360, 494)	-49	(-142, 44)	**0.298**	-986-888
Higher fibre less healthy cereals (kJ)	324	(268, 392)	290	(240, 350)	4.4	(-79, 88)	**0.917**	-838-847
Lower fibre less healthy cereals (kJ)	315	(254, 391)	394	(336, 462)	188	(98, 279)	<0.001	-724-1101
Nuts and legumes (kJ)	133	(94, 188)	176	(138, 226)	68	(12, 123)	0.018	-493-628
Fruits (kJ)	474	(415, 541)	567	(499, 644)	113	(36, 190)	0.004	-662-889
Vegetables (kJ)	90	(73, 110)	86	(69, 106)	-4.4	(-32, 23)	**0.753**	-283-274
Potatoes (kJ)	142	(112, 181)	133	(111, 159)	11	(-28, 50)	**0.589**	-384-406
Dairy (kJ)	615	(544, 696)	765	(662, 884)	240	(125, 354)	<0.001	-912-1392
Yoghurt (kJ)	190	(161, 226)	193	(159, 233)	49	(17, 80)	0.003	-269-367
Meat, fish, egg (kJ)	468	(404, 542)	760	(672, 860)	320	(235, 405)	<0.001	-537-1176
Miscellaneous (kJ)	930	(828, 1044)	830	(734, 939)	-88	(-225, 48)	**0.201**	-1462-1285

Bold = not statistically significant at $p < 0.05$. Abbreviations: CI, confidence interval; diff, difference; LOA, limits of agreement. [a] FFQ1 or FFQ2 was randomly chosen for each participant. [b] Geometric mean. [c] Paired t-test. [d] Bland-Altman limits of agreement.

Table 5.11: Mean daily intake, mean difference and limits of agreement for food groups (in grams) of relevance to the gut microbiota according to WDR and FV-FFQ in 5-year old children (n = 100).

Food group	WDR		FV-FFQ [a]		FV-FFQ [a] vs WDR			
	Mean [b]	(95% CI)	Mean [b]	(95% CI)	Mean diff	(95% CI)	p-value [c]	LOA [d]
Higher fibre more healthy cereals (g)	75	(65, 88)	72	(62, 85)	1.7	(-11, 14)	**0.794**	-124, 128
Lower fibre more healthy cereals (g)	51	(43, 60)	81[6]	(70, 94)	39	(25, 53)	<0.001	-105, 183
Higher fibre less healthy cereals (g)	19	(16, 23)	16	(13, 20)	0.3	(-5.4, 5.9)	**0.922**	-57, 57
Lower fibre less healthy cereals (g)	19	(15, 24)	28	(23, 33)	17	(10, 24)	<0.001	-52, 86
Nuts and legumes (g)	10	(7, 14)	16	(12, 21)	11	(5.3, 16)	<0.001	-43, 64
Fruits (g)	178	(155, 204)	214	(188, 243)	47	(18, 76)	0.002	-246, 339
Vegetables (g)	70	(59, 84)	55	(43, 68)	-5.9	(-21, 9.0)	**0.433**	-156, 144
Potatoes (g)	34	(28, 42)	19	(16, 22)	-9.1	(-15, -3.4)	0.002	-67, 49
Dairy (g)	139	(118, 163)	169	(138, 207)	70	(37, 103)	<0.001	-262, 403
Yoghurt (g)	61	(51, 71)	55	(45, 66)	8.6	(-0.7, 18)	**0.068**	-85, 102
Meat, fish, egg (g)	68	(60, 78)	89	(78, 100)	24	(15, 32)	<0.001	-64, 111
Miscellaneous (g)	85	(71, 100)	100	(86, 117)	15	(-2.7, 33)	**0.094**	-167, 197

Bold = not statistically significant at $p < 0.05$. Abbreviations: CI, confidence interval; diff, difference; LOA, limits of agreement. [a] FFQ1 or FFQ2 was randomly chosen for each participant. [b] Geometric mean. [c] Paired t-test. [d] Bland-Altman limits of agreement.

Table 5.12: Food group (energy contribution) correlations between the FV-FFQ and WDR (n = 100), and reproducibility correlations (n = 99) in 5-year old children, as published in Leong et al. (2018b).

	Relative validity [a]	Reproducibility [b]
Food group	**FV-FFQ [c] vs WDR**	**FV-FFQ1 vs FV-FFQ2**
Higher fibre more healthy cereals (kJ)	0.37	0.84
Lower fibre more healthy cereals (kJ)	0.35	0.91
Higher fibre less healthy cereals (kJ)	0.38	0.80
Lower fibre less healthy cereals (kJ)	0.31	0.82
Nuts and legumes cereals (kJ)	0.45	0.69
Fruits (kJ)	0.42	0.83
Vegetables (kJ)	0.28	0.78
Potatoes (kJ)	0.51	0.57
Dairy (kJ)	0.50	0.89
Yoghurt (kJ)	0.54	0.81
Meat, fish, egg (kJ)	0.56	0.83
Miscellaneous (kJ)	0.28	0.84

[a] Spearman's correlation coefficients. [b] Intra-class correlation coefficients. [c] FV-FFQ1 or FV-FFQ2 was randomly chosen for each participant.

Table 5.13: Food group (in grams) correlations between the FV-FFQ and WDR (n = 100), and reproducibility correlations (n = 99) in 5-year old children.

	Relative validity [a]	Reproducibility [b]
Food group	**FV-FFQ [c] vs WDR**	**FV-FFQ1 vs FV-FFQ2**
Higher fibre more healthy cereals (g)	0.40	0.83
Lower fibre more healthy cereals (g)	0.39	0.88
Higher fibre less healthy cereals (g)	0.37	0.79
Lower fibre less healthy cereals (g)	0.28	0.81
Nuts and legumes (g)	0.38	0.70
Fruits (g)	0.33	0.80
Vegetables (g)	0.29	0.86
Potatoes (g)	0.48	0.63
Dairy (g)	0.62	0.93
Yoghurt (g)	0.54	0.81
Meat, fish, egg (g)	0.63	0.82
Miscellaneous (g)	0.53	0.84

[a] Spearman's correlation coefficients. [b] Intra-class correlation coefficients. [c] FV-FFQ1 or FV-FFQ2 was randomly chosen for each participant.

Table 5.14: Food group (energy contribution) cross-classifications between FV-FFQ and WDR quartiles in 5-year old children (n = 100), as published in Leong et al. (2018b).

Food group	Cross-classification (FV-FFQ [a] vs WDR)			
	% Correctly classified [b]	% Correct & adjacent [c]	% Grossly misclassified [d]	% Correct extremes [e]
Chance	*25%*	*62.5%*	*12.5%*	*12.5%*
Higher fibre more healthy cereals (kJ)	36	75	5	21
Lower fibre more healthy cereals (kJ)	37	72	5	20
Higher fibre less healthy cereals (kJ)	33	78	7	22
Lower fibre less healthy cereals (kJ)	30	69	6	20
Nuts and legumes (kJ)	38	78	4	25
Fruits (kJ)	39	74	5	23
Vegetables (kJ)	28	69	5	17
Potatoes (kJ)	41	82	3	24
Dairy (kJ)	42	84	6	24
Yoghurt (kJ)	47	83	4	29
Meat, fish, egg (kJ)	51	79	2	29
Miscellaneous (kJ)	32	71	9	20

[a] FFQ1 or FFQ2 was randomly chosen for each participant. [b] % Correctly classified = percentage of children with WDR and FFQ intakes in the same quartile. [c] % Correct and adjacent = percentage of children with WDR and FFQ intakes in the same and adjacent quartiles. [d] % Grossly misclassified = percentage of children with WDR intakes in the highest quartile and FFQ intakes in the lowest quartile and vice versa. [e] % Correctly classified to extreme quartiles = percentage of children with WDR and FFQ intakes correctly classified to the lowest and highest quartiles.

Table 5.15: Food group (in grams) cross-classifications between FV-FFQ and WDR quartiles in 5-year old children (n = 100).

Food group	Cross-classification (FV-FFQ [a] vs WDR)			
	% Correctly classified [b]	% Correct & adjacent [c]	% Grossly misclassified [d]	% Correct extremes [e]
Chance	*25%*	*62.5%*	*12.5%*	*12.5%*
Higher fibre more healthy cereals (g)	35	74	7	20
Lower fibre more healthy cereals (g)	37	79	6	23
Higher fibre less healthy cereals (g)	37	75	5	22
Lower fibre less healthy cereals (g)	33	70	6	21
Nuts and legumes (g)	38	78	8	22
Fruits (g)	40	71	7	20
Vegetables (g)	30	70	8	18
Potatoes (g)	40	83	3	25
Dairy (g)	53	86	3	31
Yoghurt (g)	51	83	4	30
Meat, fish, egg (g)	46	84	2	30
Miscellaneous (g)	41	85	4	25

[a] FFQ1 or FFQ2 was randomly chosen for each participant. [b] % Correctly classified = percentage of children with WDR and FFQ intakes in the same quartile. [c] % Correct and adjacent = percentage of children with WDR and FFQ intakes in the same and adjacent quartiles. [d] % Grossly misclassified = percentage of children with WDR intakes in the highest quartile and FFQ intakes in the lowest quartile and vice versa. [e] % Correctly classified to extreme quartiles = percentage of children with WDR and FFQ intakes correctly classified to the lowest and highest quartiles.

Trends for the actual values for surrogate categories show the expected increase across the FFQ quartiles for all food groups (all $p \leq 0.028$) (**Table 5.16**). The FV-FFQ clearly differentiated between the first and fourth quartile for all food groups (all differences between the first and fourth quartile $p \leq 0.033$). The actual values for surrogate categories for the amount eaten in grams from each food group can be found in **Table 5.17**.

However, the limits of agreement were wide for all food group intakes (**Table 5.10** and **Table 5.11**). Bland-Altman plots show some bias in the food group estimates from the FV-FFQ with the FFQ having less agreement with the WDR at higher intakes (**Figure 5.3**).

Table 5.16: Ability of the FV-FFQ to differentiate between quartiles of WDR food group intake (energy contribution), determined using actual values for surrogate categories (n = 100), as published in Leong et al. (2018b).

Food group	***Quartiles defined by*** [a]	**Mean WDR intake**				**p-value for trend** [b]	**p-value for Q1 vs Q4** [c]
		Q1	***Q2***	***Q3***	***Q4***		
Higher fibre more healthy cereals (kJ)	FV-FFQ	719	1051	1070	1347	**<0.001**	**<0.001**
	WDR	357	780	1161	1890		
Lower fibre more healthy cereals (kJ)	FV-FFQ	431	556	668	781	**0.002**	**0.004**
	WDR	172	422	642	1199		
Higher fibre less healthy cereals (kJ)	FV-FFQ	235	329	496	514	**0.002**	**0.009**
	WDR	30	183	411	934		
Lower fibre less healthy cereals (kJ)	FV-FFQ	152	334	393	386	**0.028**	**0.033**
	WDR	0	120	314	856		
Nuts and legumes (kJ)	FV-FFQ	76	120	236	424	**<0.001**	**<0.001**
	WDR	0	55	199	601		
Fruits (kJ)	FV-FFQ	388	576	550	712	**0.001**	**<0.001**
	WDR	224	439	615	949		
Vegetables (kJ)	FV-FFQ	72	141	152	156	**0.016**	**0.015**
	WDR	28	70	124	299		
Potatoes (kJ)	FV-FFQ	68	136	174	294	**<0.001**	**<0.001**
	WDR	1	68	165	439		
Dairy (kJ)	FV-FFQ	478	606	835	910	**<0.001**	**<0.001**
	WDR	257	519	843	1208		
Yoghurt (kJ)	FV-FFQ	65	138	220	291	**<0.001**	**<0.001**
	WDR	0	92	215	394		
Meat, fish, egg (kJ)	FV-FFQ	287	574	598	829	**<0.001**	**<0.001**
	WDR	186	388	621	1094		
Miscellaneous (kJ)	FV-FFQ	849	1021	1213	1213	**0.010**	**0.021**
	WDR	484	854	1142	1816		

Bold = significant ($p < 0.05$). [a] FFQ1 or FFQ2 was randomly chosen for each participant. [b] Significant difference in the trend across the quartiles (regression). [c] Significant differences between Q1 vs Q4 (regression).

Table 5.17: Ability of the FV-FFQ to differentiate between quartiles of WDR food group intake (grams), determined using actual values for surrogate categories (n = 100).

Food group	*Quartiles defined by* [a]	Mean WDR intake				p-value for trend [b]	p-value for Q1 vs Q4 [c]
		Q1	*Q2*	*Q3*	*Q4*		
Higher fibre more healthy cereals (g)	FV-FFQ	65	89	102	113	**0.001**	**0.001**
	WDR	31	68	106	165		
Lower fibre more healthy cereals (g)	FV-FFQ	44	53	77	88	**<0.001**	**0.002**
	WDR	16	42	72	132		
Higher fibre less healthy cereals (g)	FV-FFQ	13	24	25	30	**0.025**	**0.021**
	WDR	2	11	25	57		
Lower fibre less healthy cereals (g)	FV-FFQ	14	16	22	27	**0.057**	0.085
	WDR	0	7	19	54		
Nuts and legumes (g)	FV-FFQ	9	11	21	25	**0.001**	**0.004**
	WDR	0	4	16	46		
Fruits (g)	FV-FFQ	165	194	205	266	**0.002**	**0.002**
	WDR	86	167	228	350		
Vegetables (g)	FV-FFQ	58	101	93	104	**0.017**	**0.008**
	WDR	23	65	94	174		
Potatoes (g)	FV-FFQ	16	28	38	54	**<0.001**	**<0.001**
	WDR	0	20	40	76		
Dairy (g)	FV-FFQ	84	139	194	288	**<0.001**	**<0.001**
	WDR	45	121	196	343		
Yoghurt (g)	FV-FFQ	21	45	66	91	**<0.001**	**<0.001**
	WDR	0	30	70	120		
Meat, fish, egg (g)	FV-FFQ	46	74	85	113	**<0.001**	**<0.001**
	WDR	30	60	89	139		
Miscellaneous (g)	FV-FFQ	60	84	141	179	**<0.001**	**<0.001**
	WDR	32	69	116	248		

Bold = significant ($p < 0.05$). [a] FFQ1 or FFQ2 was randomly chosen for each participant. [b] Significant difference in the trend across the quartiles (regression). [c] Significant differences between Q1 vs Q4 (regression).

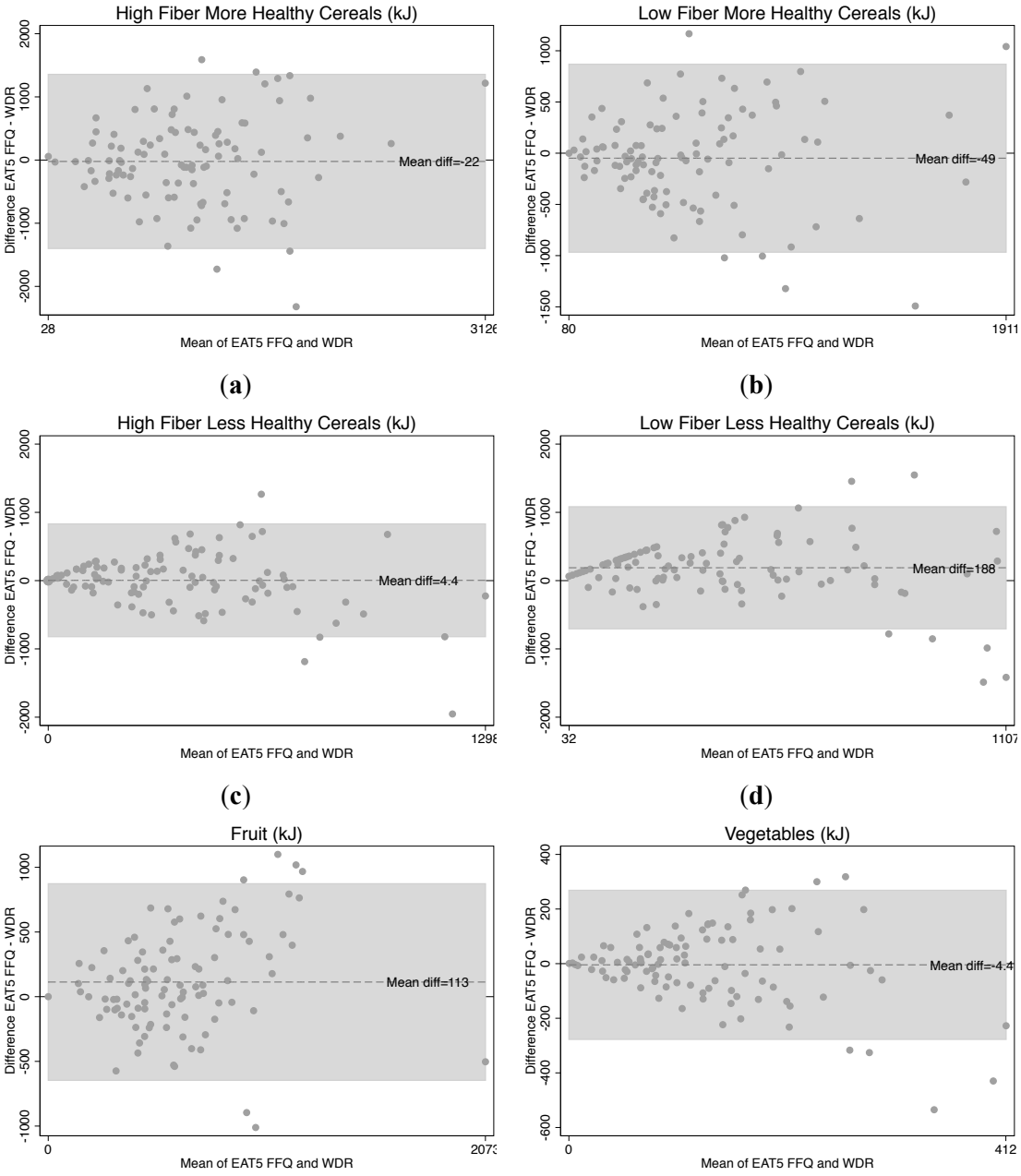


Figure 5.3: Bland-Altman plots of food group intakes (energy contribution) from a randomly chosen FV-FFQ1 or FV-FFQ2 and WDR.

(a) Higher fibre more healthy cereals; (b) Lower fibre more healthy cereals; (c) Higher fibre less healthy cereals; (d) Lower fibre less healthy cereals; (e) Fruit; and (f) Vegetables.

5.4 Discussion

Data from the FV-adjusted EAT5 FFQ were closer to the WDR data than those from the crude unadjusted EAT5 FFQ. Therefore, this discussion focuses on the FV-adjusted EAT5 FFQ, referred to as the EAT5 FFQ from now on.

The EAT5 FFQ was designed to measure intakes of nutrients and food groups of relevance to the gut microbiota in 5-year old children and showed acceptable validity, and very high reproducibility, for these over a 4-week period. The FFQ provided good estimates of mean intakes of carbohydrate, dietary fibre and total NSP intake, although it tended to overestimate energy (by 14%) and soluble NSP (32%), and underestimate insoluble NSP (by 18%) compared to the WDR; ranked most intakes acceptably (measured by correlation); and was able to differentiate well between categories of intake. Specifically, the EAT5 FFQ assigned children to correct quartiles of intake well, with very few children being grossly misclassified into the opposite quartile of intake, and was able to clearly differentiate between low and high intakes identified in the WDR.

It is difficult to compare these results directly with the literature given that no FFQs have been validated to specifically measure nutrients and foods of relevance to the gut microbiota in children. However, a number of validation studies have measured intake of energy and nutrients such as carbohydrate and dietary fibre. The correlations observed in the current study for these nutrients were within the range of those obtained in previous FFQ validation studies in young children (Blum et al., 1999; Matos et al., 2012; Parrish et al., 2003; Rodriguez et al., 2017; Stein et al., 1992; Vioque et al., 2016). Adjusting the WDR data to better reflect usual intake (using MSM) resulted in a small improvement in correlation values.

The cross-classification results for energy, carbohydrate and dietary fibre were similar to (Moghames et al., 2016) and better than (Rodriguez et al., 2017) other FFQ validation studies in young children that have reported these data. Unfortunately, the food group correlations and cross-classifications cannot be compared to previous studies because food groupings depend on the nutrients of focus for the specific validation study, so are different for different studies. However, the FFQ appeared to perform well in this context, with correlation values and gross-misclassification values for the food groups being comparable to those obtained for the nutrients in this study.

'Absolute values for surrogate categories' showed the expected stepwise increase for all nutrients and food groups and suggested that the EAT5 FFQ clearly differentiated between highest and lowest quartiles for all nutrients and food groups tested. The 'absolute values for surrogate categories' approach was developed by Willett (1998), and although it has not been commonly reported, it has been used in the validation of calcium intakes in children (Taylor & Goulding, 1998) and iron intakes in adults (Heath, Skeaff, & Gibson, 1999). It is a useful validation tool as it indicates the extent to which an FFQ is able to differentiate between broad categories of intake, as is often required in epidemiological studies.

The reproducibility of the EAT5 FFQ was consistently high, with mean correlations for reproducibility for nutrients of 0.83 and food groups of 0.80. This was higher than the range of 0.26 to 0.78 that was found for the same nutrients in other FFQ validation studies (Moghames et al., 2016; Rodriguez et al., 2017; Vioque et al., 2016). Bland-Altman plots for nutrients showed little bias in the EAT5 FFQ with a good scatter seen in the plots, but there was some bias for the food groups, particularly at higher intakes. As expected of an FFQ, the EAT5 FFQ had wide limits of agreement for nutrients and food groups suggesting that it is not appropriate for determining nutrient or food intake in individuals.

This study has several strengths. First, the EAT5 FFQ is the first FFQ validated to measure total, soluble and insoluble NSP intakes, and food groups of relevance to the gut microbiota in children. The only other studies that have validated FFQs for nutrients of relevance to the gut microbiota (i.e. intake of NSPs (Emmett, Symes, Braddon, & Heaton, 1992), fibre (Healey et al., 2016; Reeves, Winkler, & Eakin, 2015), inulin and oligosaccharides (Dunn et al., 2011)) have been in adults. Interestingly, the current study had lower correlations (i.e. they were 'acceptable') than the correlations that were obtained in an adult study validating intake of NSPs (which were 'acceptable' to 'good') (Emmett et al., 1992). A possible explanation may be that parental proxy reporting acts as an additional layer of potential error in assessing diets in children. However, the studies validating dietary fibre intake in adults (Healey et al., 2016; Reeves et al., 2015) also used another questionnaire as the reference method, rather than a diet record. This means that they did not use a widely accepted method of dietary assessment with different errors to those of the FFQ as a reference method, as is recommended (Cade et al., 2002), and as was used in the current study. Second, the aim

was to validate several carbohydrate food groups of interest defined by their fibre content and overall healthiness using strict criteria. By contrast, most previous studies have combined all carbohydrate containing foods into a single group such as cereal or grains (Huybrechts, De Backer, De Bacquer, Maes, & De Henauw, 2009; Mills et al., 2015) with most of the 'less healthy' carbohydrate foods appearing in the 'snacks' food group, even if they contained fibre (Flood et al., 2013). Third, a non-consecutive 3-day WDR was used as the reference method. Many other FFQ validation studies in children have used 24-hour recalls (Blum et al., 1999; Matos et al., 2012; Moghames et al., 2016; Parrish et al., 2003; Rodriguez et al., 2017; Vioque et al., 2016) or estimated diet records (Vereecken et al., 2010), both of which have similar errors to an FFQ with potential for memory lapses and errors in portion size estimation. Finally, cross-check questions were used for the fruit and vegetable sections. Using the cross-check questions improved the performance of the EAT5 FFQ (as compared to the unadjusted values). This is particularly important, as the EAT5 FFQ was developed to look at nutrients of relevance to the gut microbiota and hence has a large number of fruit and vegetable questions. Fruit and vegetable food groups have been shown to be commonly over-reported in other FFQ validation studies (Calvert, Cade, Barrett, & Woodhouse, 1997).

The current study has some limitations. First, the ethnicity of the participants is not representative of the New Zealand population as a whole, with a higher proportion of New Zealand Europeans, and an over-representation of participants from the lower and middle deciles of household deprivation. Second, the FFQ was administered only to the primary caregiver, and as the children were 5 years of age, they would be attending school, so were not with their parents at all times. However, in an effort to overcome this common limitation, parents were asked to report any food and amount eaten if someone else provided their child with food, and the child attended the appointment with their parent so was available for parents to clarify their answers. Third, it was only possible to test the ability of the FFQ to measure intake of fibre and total, soluble and insoluble NSPs, not intake of smaller fibre fractions such as pectin, cellulose, hemicellulose and klason lignin that may also impact on the gut microbiota. This was because these fractions are not measured and reported in the New Zealand food composition database, FOODfiles, and the literature was not sufficient to provide reliable data for all 1010 foods consumed in the WDRs.

This study shows that the EAT5 FFQ has acceptable validity when compared with a 3-day WDR and has very good reproducibility when measured over 4 weeks. It is suitable for assessing mean absolute intake of carbohydrate, dietary fibre, and total NSP. However, it is unable to assess intakes of dietary fibre fractions such as resistant starch and oligosaccharides. The EAT5 FFQ is able to rank the diets of young children adequately, and to correctly assign low and high intakes of nutrients and food groups of relevance to the gut microbiota. This Chapter provides evidence that the EAT5 FFQ is an appropriate tool for measuring the intake of nutrients and food groups of relevance to the gut microbiota in 5-year olds, and supports its use in the POI study reported in **Chapter 7**.

6 POI Methods

This chapter outlines the POI study more generally to provide background for the specific methods that are relevant to this thesis. POI was a 2-year intervention (antenatal to 2 years of age), with follow-up measurements undertaken at 3.5 and 5 years of age (no intervention during this time). Further information on the POI study can be found in the protocol papers for POI (Taylor et al., 2011; Taylor et al., 2016), and the main outcomes paper from the end of the two year randomised controlled trial (RCT) (Taylor et al., 2017a) and the 5 year follow-up (Taylor et al., 2018).

6.1 Study design

The Prevention of Overweight in Infancy (POI) intervention was a four-arm RCT that recruited 802 families from Dunedin, NZ. The main aim of the POI study was to investigate whether a conventional approach promoting healthy eating and activity or a more indirect approach targeting sleep during late pregnancy and early childhood would result in lower body mass index (BMI) at 2 years of age compared to standard care (Taylor et al., 2017a). The POI study had ethical approval from the Lower South Regional Ethics Committee (LRS 08/12/063), and is registered under the clinical trials registration NCT00892983.

The aim of the POI follow-up was to determine the extent to which the POI interventions on infant sleep, feeding, diet and physical activity in the first 2 years of life reduces BMI at 3.5 and 5 years of age. All participants from the original POI intervention were invited to participate in the follow-up study (unless they had specifically requested no further follow-up). The POI follow-up study had approval from the University of Otago Human Ethics Committee (12/ 274). Written informed consent was obtained from adult participants during pregnancy (intervention) and before commencing the first follow-up visit at 3.5 years of age (follow-up).

6.1.1 Participants

802 pregnant women were recruited between May 2009 and November 2010 from the Queen Mary Maternity Centre (Dunedin Hospital, Dunedin, NZ). Queen Mary Maternity Centre is the only birthing unit in the city of Dunedin and provides primary, secondary and tertiary maternity services for more than 97% of births in Dunedin city. Inclusion criteria were: mothers who were booked into Queen Mary Maternity Centre before 34 weeks' gestation, intended to live in Dunedin for the next two years, were able to communicate in English or Te Reo Māori (the language of the indigenous people of New Zealand), and were 16 years of age or older. Exclusion criteria were: infants born before 36.5 weeks' gestation, or infants having a congenital abnormality or physical or intellectual disability that was likely to affect their feeding, physical activity or growth. The lead maternity carer (usually a midwife) provided study information to women with planned home births (< 3%). Participants were sent a letter of invitation to participate in the study with an opt-out option at 28 weeks' gestation.

6.1.2 Sample size

For the POI study, the power calculation was conducted based on sufficient power to detect a 0.5 kg/m^2 difference in BMI (primary outcome) at two years of age. For the POI follow-up, the study was powered on detecting a difference in BMI of 0.3 kg/m^2 at 5 years of age. For the microbiota component, which is the main outcome in this thesis, sample size was not calculated as it was a secondary outcome only.

6.1.3 Randomisation

Participants were randomised to 4 groups via stratified block allocation (with a block size of 12) with stratification for parity (first child or subsequent child) and socioeconomic status (low, medium or high). Allocation was concealed using opaque pre-sealed envelopes. Out of the 1458 that were eligible to participate in the study, 611 declined to participate and 45 were excluded after birth. A total of 802 participants were assigned to one of the 4 arms in the study: Control (n = 209), 'Food, activity and breastfeeding' FAB (n = 205), Combination of FAB and Sleep (n = 196) and Sleep (n = 192) (**Figure 6.1**).

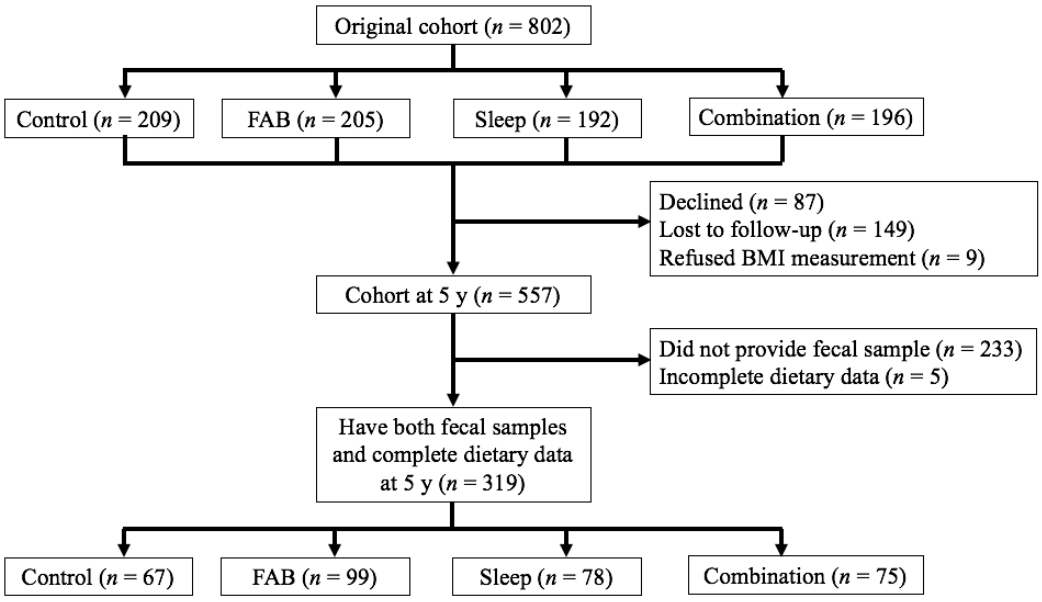


Figure 6.1: Flow of participants through the study for the microbiota component of the POI study.

FAB, group receiving information on food, activity and breastfeeding; Sleep, group receiving sleep intervention training; Combination group receiving both FAB information and sleep intervention.

6.2 Study groups and intervention

Similar to the BLISS study (section 3.2), all four groups received routine midwifery care until 6 weeks of age and Well Child care after this time. Well Child Tamariki Ora is a free health care programme for all children in NZ under 5 years of age. This programme involves free home and clinic visits by trained community nurses who provide advice that covers child growth and development, and checks on oral health, vision, hearing and overall health and development (Ministry of Health, 2017).

6.2.1 Control

The Control group received Well Child Tamariki Ora routine care (as described in section 3.2), with no additional intervention.

6.2.2 Intervention groups

The intervention groups received Well Child Tamariki Ora routine care (as described in section 3.2), and additional parent contacts from the POI study team for support and education from birth to 2 years of age. The POI study team that delivered the intervention included specially trained research nurses and lactation consultants. A brief description of the interventions is summarised in **Table 6.1**.

6.3 Infant, maternal and family characteristics at baseline

Demographic data such as maternal age, ethnicity, education, income and parity were collected at baseline (19 - 39 weeks' gestation). New Zealand Index of Deprivation (NZDep) score for each household (Atkinson et al., 2014) was determined using the participant's current address providing an overall index of the level of household deprivation where 1 refers to low deprivation and 10 to high deprivation. Maternal pre-pregnancy BMI was calculated through self-reported weight at baseline and height measured when the infant was 6 months of age. Infant sex, birth weight, and gestational age at birth were accessed through hospital records.

Table 6.1: Summary of POI intervention groups.

FAB	Sleep	Combination
Third trimester of pregnancy to 2 years		
8 (minimum) additional contacts	2 (minimum) additional contacts	9 (minimum) additional contacts
• Education and support around breastfeeding, food and activity. • At 34 - 35 weeks' gestation: discuss breastfeeding and develop a breastfeeding plan. • First week and 4 months: support from lactation consultant around exclusive breastfeeding to 6 months, breastfeeding to 12 months and introduction of solids following the 'you provide, they decide' model. • 3, 9 and 18 months: group physical activities to encourage 'tummy time', outdoor play and family activity. To limit television viewing prior to 2 years. • 7, 12 and 18 months: food-based education sessions.	• Education and support around developing appropriate sleep habits from early in life. • At 34 - 35 weeks' gestation: outline normal sleep and techniques to prevent the development of sleep problems. • 3 weeks: home visit for support from research nurse and information booklet about developing healthy sleep patterns • 6 months to 2 years: optional additional support from research nurse and Sleep research team given to those parents who requested assistance because they perceived their child had a sleep problem.	• Receive all the FAB and Sleep interventions.

Abbreviation: FAB, Food, activity and breastfeeding.

6.4 Anthropometric measurements

Infant length and weight measures at birth were obtained from hospital records. At 6, 12, 18, and 24 months, 3.5 and 5 years, length/height and weight measures were taken by trained researchers following the World Health Organization (WHO) protocols (de Onis et al., 2004). Weight-for-age z-scores and BMI-for-age z-scores were calculated using the WHO growth standards (WHO Multicentre Growth Reference Study Group, 2006), with overweight defined as a BMI z-score ≥ 1.036 to < 1.645 and obesity as ≥ 1.645 (the z-scores for the 85^{th} and 95^{th} percentiles of the POI study population).

6.5 Body composition

At 5 years of age, fat mass, fat-free mass and percentage of body fat were assessed by dual-energy x-ray absorptiometry (DXA). All DXA measurements were performed and analysed by one experienced operator with a Lunar Prodigy scanner (software package 16.0; Lunar, Madison, WI) using standard procedures. Fat mass index (FMI) was calculated as fat mass divided by height in metres squared. The FMI z-scores were then calculated for the study sample (Weber, Moore, Leonard, & Zemel, 2013).

6.6 Dietary assessment

6.6.1 Breastfeeding and introduction to solids

Mothers completed feeding questionnaires by telephone every 4 weeks when their infants were 3 to 27 weeks of age. This allowed the determination of exclusive breastfeeding (no other liquids or solids since birth) and full breastfeeding (no other liquids or solids in the past 48 hours) status *to the nearest day* as well as the timing of the introduction of solids (rather than by recollection at 6 - 12 months of age which is more typical). In addition, at 12, 18, 24, 42 and 60 months, mothers were asked if they were still breastfeeding (if they were breastfeeding in the previous questionnaire) and if not, the age when they stopped, to determine length of 'any' breastfeeding.

6.6.2 Food frequency questionnaires

Dietary intake was assessed by a validated food frequency questionnaire (FFQ) (Watson et al., 2015) at 12 and 24 months for both frequency and amounts of 91 food items consumed over the past month (Fangupo et al., 2015). At 5 years, dietary intake was assessed using the validated EAT5 FFQ (validation study detailed in **Chapter 5**) for both frequencies and amounts (123 food items) of food eaten over the past month (Leong et al., 2018b). The POI5 FFQ refers to the EAT5 FFQ in this chapter.

6.6.3 Entering of FFQ

The POI study team administered the FFQs, and the Candidate was responsible for entering and checking all of the POI5 FFQ (n = 546) data from the hardcopies into an online database (ffq.otago.ac.nz). A codebook was developed by the Candidate to enter the data to ensure data consistency (**Appendix I**). This codebook had rules such

as amounts to enter if the amount was not specified. For example, if the 'Mince & patties' amount was entered as '1 patty' rather than in mL (as requested), a figure of '50 mL' was used for that entry based on unit values in grams and density of the product in Kai-culator (dietary assessment software as mentioned in **Chapter 5**). This codebook was developed with consultation from a senior dietitian who has many years' experience in coding dietary data, especially in national nutrition surveys. After entering the data, the output was checked by the Candidate. The Candidate followed the protocol described in **Appendix J** for data entry checking.

6.7 Dietary analysis

6.7.1 Nutrient analysis

Data on the key nutrients and food components of interest (energy and dietary fibre fractions) for the POI5 FFQ were sourced from developed nutrient lines (**Chapter 5**). The values for amount consumed per week were exported from the online database to be used in the statistical analysis. This output was checked for accuracy by the Candidate as described in **Chapter 5**.

6.7.2 Food group analysis

The Candidate developed 12 food groups that were considered, based on the literature, to have the potential to be relevant to the gut microbiota as well as general health. Details of the 12 food groups can be found in section 5.2.8. In short, the 123 FFQ food items were allocated to the 12 food groups. The Candidate used the Stata 13 software to combine the nutrient output from the individual FFQ food items (sum of frequencies and amounts) into the 12 food groups. This procedure is similar to that described in section 5.2.8.

6.8 Gut microbiota

Details of faecal collection and DNA extraction are similar to those detailed in section 3.6.1 and data preparation was carried out by the same personnel as in section 3.6.2. Briefly, a faecal sample was collected from participants at 5 years of age (approximately 5 g). The sample was then stored in the home freezer (-18°C) in study-provided freezer containers before collection and delivery to the Department of Microbiology and Immunology, University of Otago, where they were stored at -80°C until DNA extraction. Amplification of the 16S rRNA gene V4 region, library

preparation and sequencing were carried out at Argonne National Laboratories (University of Chicago) using 2 x 250 base paired-end reads on an Illumina MiSeq instrument. The data preparation was carried out by Blair Lawley from the Department of Microbiology and Immunology, University of Otago. The Sequences were analysed using a combination of QIIME version 1.9.1 (Caporaso et al., 2010) and vsearch version 1.9.5 (Rognes et al., 2016). Taxonomy classifications were made using the SILVA version 123 database (Quast et al., 2013). Raw sequence data can be publicly assessed at NCBI (Accession Number: PRJNA528813).

6.9 Statistical analysis

All relevant statistical analyses are discussed in **Chapter 7**.

7 Describing children's gut microbiota profiles

This chapter uses compositional principal component analysis to describe children's gut microbiota profiles and relate the profiles to diet and body composition. The statistical analysis plan and lessons learnt are included in **Appendix N**.

This chapter uses the following data collected during the Prevention of Overweight in Infancy (POI) study:

- Maternal Baseline Questionnaires (collected during pregnancy).
- Feeding Questionnaires (collected between 3 to 27 weeks).
- Gut microbiota data from faecal samples (collected at 5 years).
- Validated food frequency questionnaire (FFQ) (collected at 5 years).

Chapter highlights:

- Use of compositional principal component analysis (PCA) to describe children's gut microbiota profiles. PCA is commonly used to describe dietary patterns in the health literature and principal co-ordinate analysis (PCoA) for gut microbiota profiles in the microbiota literature. To the Candidate's knowledge, this is the first study to apply compositional PCA to children's gut microbiota data.
- Use of gut microbiota profile scores to determine associations with both diet and body composition. To the Candidate's knowledge, this is the first study to assign scores for the different profiles to each individual participant, and use these scores to examine associations between diet and body composition in children.
- Use of an FFQ that has been validated in children of the same age group and for the dietary components considered to be of relevance to the gut microbiota. Studies in this area have to date used FFQs that have not been validated for that study population, or have not been validated for the nutrients being studied.

A paper based on this chapter is currently under review with the American Journal for Clinical Nutrition:

Leong, C., Haszard, J. J., Heath, A.-L. M., Tannock, G. W., Lawley, B., Cameron S. L., Szymlek-Gay, E. A., Gray, A. R., Taylor, B. J., Galland, B. C., Lawrence, J. A., Otal, A., Hughes, A., Taylor, R. W. (2019). Using compositional principal component analysis to describe children's gut microbiota profiles in relation to diet and body composition.

7.1 Introduction

Interest in the role the human gut microbiota may play in nutritional health is growing (De Filippo et al., 2010; Nakayama et al., 2017). For instance, dysbiosis (imbalance) of the gut microbiota community has been linked to obesity in adults (Bäckhed et al., 2012; Valdes et al., 2018). However, less is known about any relationship with childhood obesity (Taylor, 2016). Furthermore, while it is generally considered that an adult-like, stable and more diverse gut microbiota is attained by around 3 years of age (Koenig et al., 2011; Palmer et al., 2007; Tanaka & Nakayama, 2017; Yatsunenko et al., 2012), further changes appear to occur during childhood (Hollister et al., 2015). Associations between the gut microbiota and health outcomes, particularly obesity, may therefore be different in growing children to those in adults.

In recent years the structure and appropriate statistical analysis of gut microbiota data obtained from analysis of 16S rRNA sequences has come under scrutiny. As detailed in the literature review (section 2.8), due to the nature of sequencing technology these data are commonly expressed as relative abundance, which is the abundance of each taxon relative to the total abundance, expressed as a percentage or proportion. These relative abundances are compositional data because the bacterial taxa are constrained to 100% and are therefore co-dependent (i.e. if the percentage of one bacterial taxon increases, the percentage of at least one other bacterial taxon has to decrease, because by definition, the sum must always be 100%). Traditional methods of statistical analysis, such as regression modelling and t-tests, are not appropriate for such data because compositional data violate the assumptions of these analyses, in particular the assumption that variables are independent (Gloor et al., 2017; Gloor & Reid, 2016; Quinn et al., 2018; Tsilimigras & Fodor, 2016).

Nutrition researchers commonly use principal component analysis (PCA) to determine dietary patterns in an attempt to describe the complexity of human diets, rather than investigating single foods or nutrients (Hu et al., 1999). With dietary pattern analyses, all participants receive a 'score' for each pattern which indicates how strongly they adhere to that particular pattern. For example, in infants, one study identified a 'vegetables and meat' dietary pattern that was characterized by higher intakes of 'vegetables', 'meat, eggs and beans', and 'fruit', and lower intakes of 'baby and toddler foods' (Mills et al., 2015). While PCA has been used to investigate gut microbiota

profiles, the compositional nature of these data have not always been appropriately handled (Gloor & Reid, 2016), and the taxa contributing to each profile are not generally stated. Instead, results of PCA are commonly interpreted by visualizing the variables (e.g., relative abundance of genera) in the same space using a biplot (Meng et al., 2016), without undertaking further statistical analyses on the profiles generated. This could be a missed opportunity to discover whether different combinations of bacterial taxa are associated with diet or body composition.

Therefore, the aim of this study was to use compositional PCA to describe gut microbiota profiles in 5-year old New Zealand children and to explore relationships between these gut microbiota profiles and diet, BMI, and body composition.

7.2 Methods

7.2.1 Data collection

Please refer to **Chapter 6** for comprehensive details on the methods used in the POI study. In short, the POI study was a 2-year, 4-arm randomised controlled trial (RCT) that aimed to determine whether additional education and support for parents regarding infant sleep (Sleep group); food, activity and breastfeeding (FAB group); or both (Combination group), reduced excessive weight gain as compared to usual care (Control group) at the end of the intervention (2 years of age) and after 3 years of follow-up (5 years of age). A sub-set of the POI study participants (n = 319) provided faecal samples that could be used in this thesis. FFQs (validated in **Chapter 5**) and faecal samples for gut microbiota analysis (same methodology as detailed in section 3.6) were collected at 5 years of age.

7.2.2 Data preparation

In this Chapter, microbiota composition was described by i) gut microbiota profiles, ii) an alpha diversity measure, the Shannon Index which described richness and evenness (i.e. the equality of distribution of the species' frequencies) and iii) the 'Firmicutes to Bacteroidetes' (F/B) ratio (the two predominant microbial phyla in the gut that have been associated with obesity (Ley, Turnbaugh, Klein, & Gordon, 2006)). The F/B ratio was calculated by dividing the relative abundance of the Firmicutes phylum over the relative abundance of the Bacteroidetes phylum.

The gut microbiota profiles were obtained from relative abundance at the genus level, which was calculated by collapsing the raw OTU table based on seven-level taxonomy strings (kingdom_phylum_class_order_family_genus_species) obtained from the SILVA version 123 database.

For the dietary data, nutrients (carbohydrate, dietary fibre, total non-starch polysaccharides (NSP), and insoluble NSP), energy, and 9 of the 12 food group intakes in grams (i.e. 'higher fibre more healthy cereals', 'higher fibre less healthy cereals', 'lower fibre more healthy cereals', 'nuts and legumes', 'fruits', 'potatoes ', 'dairy', 'yoghurt' and 'meat, fish, eggs') were used in the analyses on the basis of having acceptable validity ($r \geq 0.30$ in comparison with diet records) (**Chapter 5**).

7.2.3 Statistical analysis

Data from all four study groups were combined for cross-sectional analyses. Data were analysed using Stata software (StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP).

Student t-tests (assuming equal variance) and chi-squared tests were used to compare baseline variables between those who provided complete microbiota and dietary data, and those who did not (Table 7.1).

The graph in **Figure 7.1** was generated using R (version 3.5.2). **Figure 7.2** was generated using GraphPad Prism (version 8; GraphPad Software, La Jolla California, USA). $P < 0.05$, two-sided for tests where this is an option, was considered to indicate statistical significance.

Gut microbiota profiles

To generate the gut microbiota profiles, compositional PCA was used. To be as robust as possible, unclassified genus-level OTUs were removed before analysis, leaving 311 genera to be analysed. Next, only genera that were present in ≥ 75% of the participants were analysed, resulting in 124 genera from the initial 311 genera for the 319 participants.

As gut microbiota data are compositional in nature (Gloor et al., 2017; Gloor & Reid, 2016; Quinn et al., 2018; Tsilimigras & Fodor, 2016), centred log-ratios of each genus-level OTU to the geometric mean of all genus-level OTUs were used. As this

transformation does not work for zero values, zeros were replaced with '0.000001' in the dataset. PCA of the centred log-ratios was carried out, and a scree plot of eigenvalues was examined to determine how many profiles were represented (three were indicated). Orthogonal rotation of the components was undertaken.

After this, each genus-level OTU had a 'loading' for each profile – simply, the loading represents the correlation with the profile so that high loadings identify the genera that characterize the profile. We described each profile by identifying those genera that had loadings of 0.15 or greater (positive or negative). Every child participant was given a score for each of the 3 gut microbiota profiles. All 124 genera were used in the calculation of each gut microbiota profile score, which is the sum of the loadings of each genus multiplied by the centred log-ratio of that genus. Hence, a child participant's score for profile 1 = (their loading of genus 1 on profile 1 x their centred log-ratio of genus 1) + (their loading of genus 2 on profile 1 x their centred log-ratio of genus 2) + … + (their loading of genus 124 on profile 1 x their centred log-ratio of genus 124). Similar calculations were undertaken for profile 2 and profile 3, enabling scores for all three profiles for every child.

To illustrate the characteristics of each profile, the median relative abundance of the 12 - 13 genera that had loadings of ≥ 0.15 (positive or negative) for each gut microbiota profile from all (n = 319) participants was calculated and displayed in **Figure 7.2**.

This method described above is different from enterotyping which will be mentioned in the discussion. In short, enterotyping assigns each participant to just 1 profile. Enterotypes for this study were developed following the steps as described in the R tutorial (https://enterotype.embl.de/enterotypes.html) by the authors of the original enterotype publication (Arumugam et al., 2011). The steps to obtain the profiles and enterotypes are illustrated in the **Figure 7.1** below.

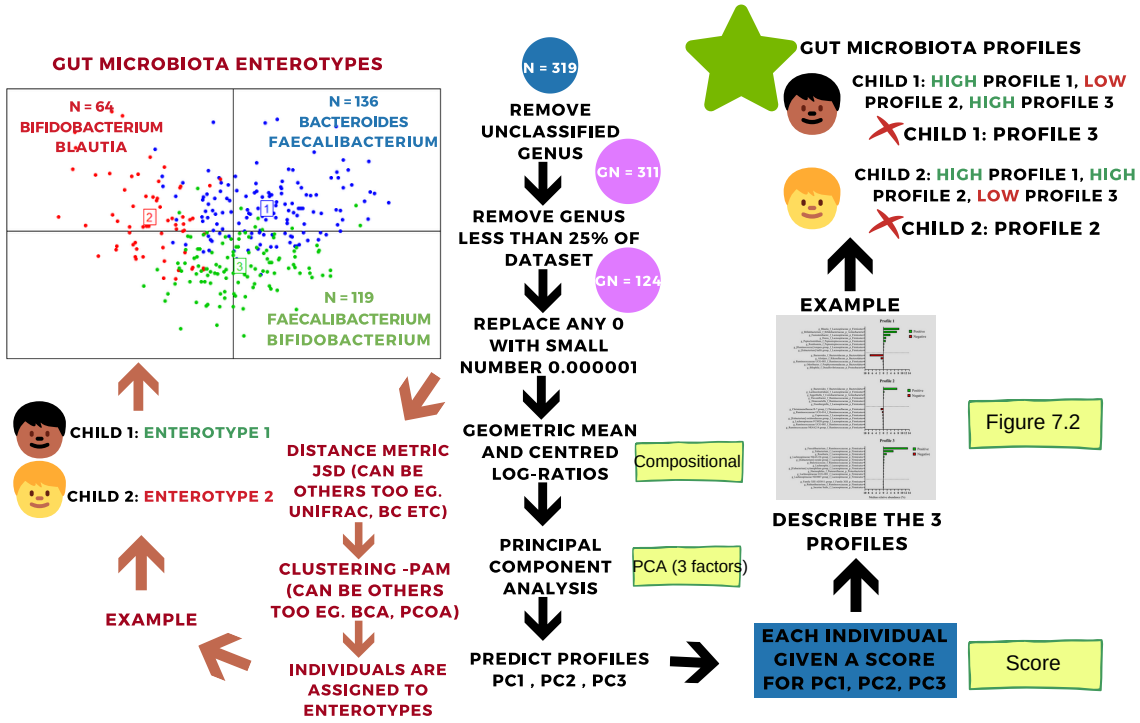


Figure 7.1: Illustration of steps taken to obtain gut microbiota profiles (in black words, towards the right) and gut microbiota enterotypes (in brown words, towards the left).

Example Figure 7.2 was developed using GraphPad Prism and gut microbiota enterotype's Example Figure developed using R. Candidate used the CANVA software for graphical design of this figure. Abbreviations: N, number of participants; GN, number of genera; PCA, principal component analysis; PC, principal component; JSD, Jensen-Shannon divergence; PAM, Partitioning around medoids clustering algorithm; BCA, between-class analysis; PCoA, principal co-ordinate analysis.

After describing the 3 gut microbiota profiles, spearman's correlations were used to initially screen for associations between diet and body composition variables and gut microbiota profiles (**Table 7.2**). Variables with the highest correlations were then explored further with four linear regression models as follows (**Table 7.3**): the first model examined the association of *demographics* reported at birth with gut microbiota profiles at 5 y of age to identify demographic predictors and possible confounders; the second model examined the cross-sectional association of *dietary components* with gut microbiota profiles (adjusted for demographic variables that were found to be possible confounders in model 1 ($p < 0.25$ (Mickey & Greenland, 1989)); the third model examined the association of *BMI z-score and FMI z-score* with gut microbiota profiles (adjusted for demographic variables that were found to be possible confounders ($p < 0.25$ (Mickey & Greenland, 1989)). The fourth model (**Table 7.4**) investigated whether

the *association between BMI z-score and the gut microbiota profile was influenced by including dietary component variables in the model and vice versa*. All models were adjusted for the four POI study groups. The results were presented as the estimate (β coefficient) and 95% CI for the difference in the predicted value of the outcome variable (i.e. gut microbiota profiles) for each 1-unit difference in the predictor variable (i.e. diet or BMI z-score or FMI z-score variables) or difference between categories (e.g., multiparous compared to primiparous). Residuals were plotted and visually assessed for homogeneity of variance and normality.

7.3 Results

7.3.1 Study population

Table 7.1 compares the demographics of the participants who provided both diet and gut microbiota data at 5 years of age (n = 319) to those who did not provide complete data (n = 483). Overall, mothers in the current analysis had a mean age of 33 y, a self-reported mean pre-pregnancy BMI of 24.6, and 85% had a post-secondary or higher education qualification. Twenty-one percent of the child participants were categorized as overweight or obese (BMI z-score ≥ 1.036). Most children (87%) were White, and were delivered by vaginal birth (71%). Around half of the children were male (53%) and had at least one sibling (55%). In total, 41% of households had low levels of deprivation, compared to the expected 30% (Atkinson et al., 2014). Compared to those who did not provide both diet and gut microbiota data, mothers of children who provided these data were generally older, better educated, had a lower pre-pregnancy BMI, and the children came from households with better socio-economic status as defined by the NZ household deprivation score (Atkinson et al., 2014).

Table 7.1: Baseline characteristics of participants.

Variables [a]	**Whole sample (n = 802)**	**Current study sample (n = 319)**	**p-value** [b] **(With vs without data)**
Child sex, n (%)			0.346
Female	391 (49)	149 (47)	
Male	411 (51)	170 (53)	
Child birth weight, (g)	3551 ± 480	3571 ± 473	0.352
Missing, n	7	1	
Child's weight category at 5 y, n (%)			0.223
Overweight/ obese [c]	125 (23)	67 (21)	
Normal	422 (77)	252 (79)	
Missing, n	255 [d]	0	
Maternal age at child's birth, (y)	31.6 ± 5.2	33.2 ± 4.5	**<0.001**
Missing, n	1	0	
Maternal ethnicity, n (%)			**0.013**
White	682 (85)	277 (87)	
Māori	46 (6)	9 (3)	
Others	74 (9)	33 (10)	
Maternal parity, n (%)			0.251
Primiparous	382 (48)	144 (45)	
Multiparous	420 (52)	175 (55)	
Maternal education, n (%)			**<0.001**
Secondary or below	193 (24)	49 (16)	
Post-secondary or higher	601 (76)	268 (84)	
Missing, n	8	2	
Maternal pre-pregnancy BMI, (kg/m^2)	25.1 ± 5.0	24.6 ± 4.5	**0.024**
Missing, n	3	0	
Household deprivation [e], n (%)			**0.005**
Low (1-3)	276 (35)	129 (41)	
Mid (4-7)	350 (44)	133 (42)	
High (8-10)	168 (21)	53 (17)	
Missing, n	8	4	
Mode of delivery, n (%)			0.736
Vaginal	562 (72)	223 (71)	
Caesarean	222 (28)	91 (29)	
Missing, n	18	5	

Bold = significant ($p < 0.05$). [a] Mean ± SD (all such values not otherwise indicated). [b] P-value compares the characteristics of the participants with diet and microbiota data (n = 319), and those without (n = 483), determined using t-tests for continuous variables and chi-square tests for categorical variables. [c] Overweight/ obese participants were those with BMI z-score ≥ 1.036 at 5 y. [d] 557 participants were retained at 5-year, and 547 participants provided both weight and height measurements. [e] Determined using the New Zealand Index of Deprivation 2013 (Atkinson et al., 2014). The Index combines 9 variables from the 2013 New Zealand National Census to provide a deprivation score for each meshblock (a geographical unit defined by Statistics New Zealand that contains on average approximately 81 people). The score reflects the extent of material and social deprivation and is used to construct deciles from 1 (low deprivation) to 10 (high deprivation).

7.3.2 Gut microbiota profiles

Figure 7.2 illustrates the three main gut microbiota profiles identified by compositional PCA using the genera detected in the children's faecal samples. Profile 1 (8.3% of variance explained) was characterized by positive loadings of dominant genera *Blautia*, *Bifidobacterium* and *Fusicatenibacter*; and negative loadings of dominant genera *Bacteroides* and *Alistipes*. Profile 2 (8.2% of variance explained) was characterized by positive loadings of dominant genera *Bacteroides*; and negative loadings of dominant genera *Christensenellaceae R-7 group* and *Ruminococcaceae UCG-014*. Profile 3 (7.5% of variance explained) was characterized by positive loadings of dominant genera, *Faecalibacterium*, *Eubacterium* and *Roseburia.*

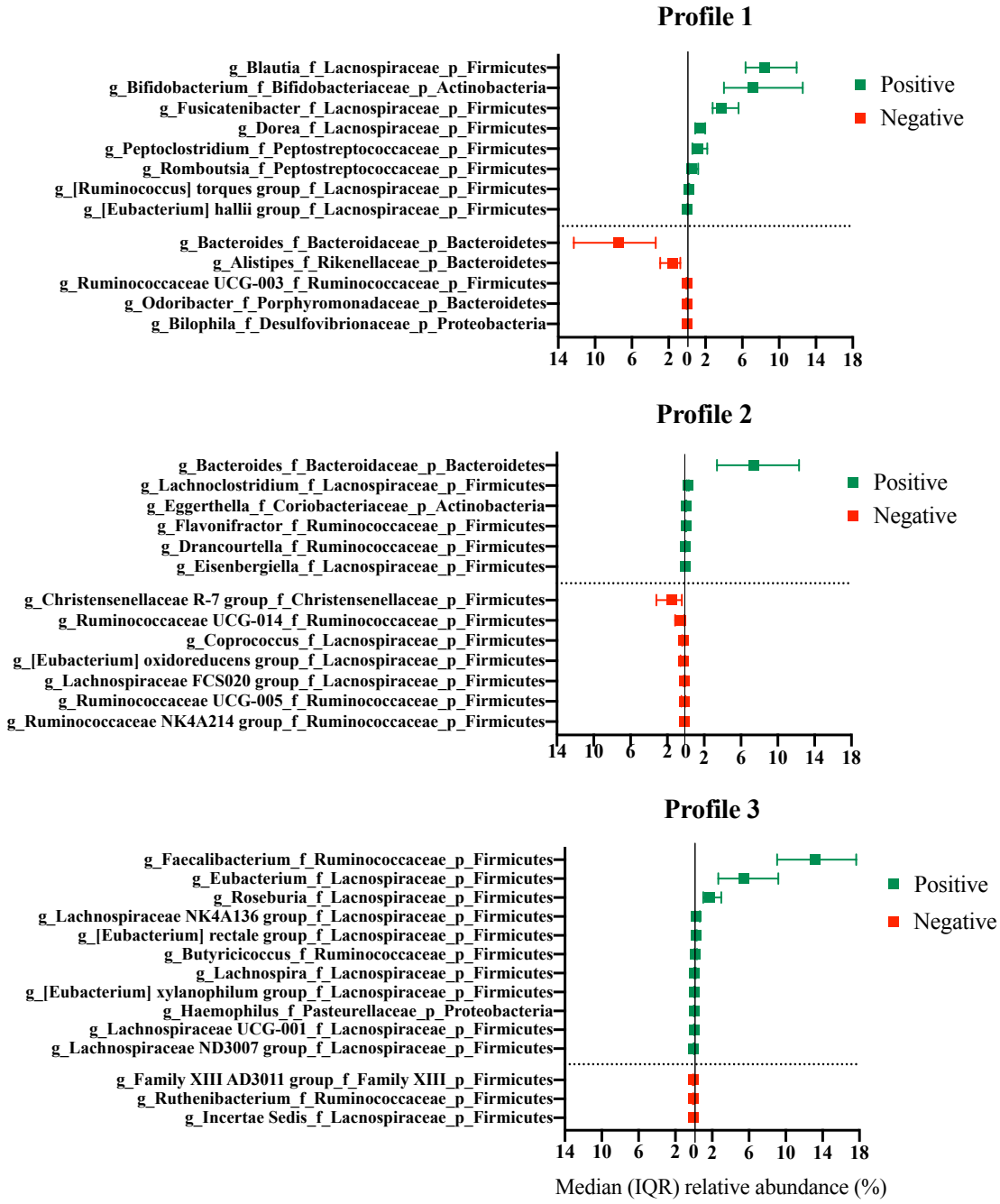


Figure 7.2: The three main gut microbiota profiles of 5-year old children, determined using principal component analyses of compositional data.

The genus (g) with its respective family (f) and phylum (p) plotted, have loadings of greater than 0.15. The bars show the median (25th to 75th percentile) relative abundance of the genus in the total study sample (n = 319). Green bars represent the positive loadings and red bars represent the negative loadings.

7.3.3 Gut microbiota associations with diet and body composition

When associations were examined between the 3 profiles and dietary and body composition variables using Spearman's correlations, profile 1 was not monotonically associated with any variable of interest (**Table 7.2**).

Table 7.2: Correlation of infant diet, nutrients at 5-year, food groups at 5-year, and body composition and BMI z-score and FMI z-score at 5-year; with Shannon index and gut microbiota profiles at 5-year.

Variables	Mean (SD)	Shannon index [b]	Profile 1	Profile 2	Profile 3
Infant diet					
Exclusively breastfed*, wk*	14 (9.7)	0.02	-0.04	0.01	0.05
Any breastfeeding*, wk*	52 (39)	-0.03	0.00	0.09	0.01
Age at introduction of solids [c]*, wk*	22 (3.4)	-0.04	0.05	-0.05	-0.04
Nutrients at 5-year (per day basis)					
Energy, *kJ*	6253 (1727)	0.03	0.03	0.02	0.07
Carbohydrate, *g*	184 (52)	0.05	-0.01	-0.04	0.05
Fat, *g*	56 (19)	-0.01	0.06	0.07	0.05
Protein, *g*	61 (19)	0.04	0.07	0.05	0.12*
Total dietary fibre, *g*	19 (6.1)	0.09	-0.03	-0.07	0.15*
Total NSP, *g*	15 (4.9)	0.08	-0.03	-0.06	0.15*
Insoluble NSP, *g*	7.2 (2.4)	0.08	-0.03	-0.07	0.13*
Food groups at 5-year (per day basis)					
Higher fibre more healthy cereals, *g*	97 (70)	0.11*	-0.05	-0.07	0.07
Lower fibre more healthy cereals, *g*	107 (78)	-0.11*	0.04	0.08	0.02
Higher fibre less healthy cereals, *g*	21 (19)	-0.02	0.03	-0.05	0.06
Nuts, seeds, legumes, *g*	24 (26)	0.01	-0.03	-0.02	0.11*
Fruits, *g*	231 (132)	0.08	-0.09	-0.06	0.09
Potatoes, *g*	16 (12)	-0.09	0.06	0.08	-0.07
Dairy, *g*	278 (260)	0.06	0.01	0.02	0.00
Yoghurt, *g*	67 (48)	0.13*	-0.05	-0.01	0.06
Meat, fish, poultry, *g*	85 (42)	0.00	0.05	0.05	0.15*
Body composition, BMI z-score and FMI z-score at 5-year					
DXA fat [d], *%*	16 (4.9)	0.04	0.03	-0.06	-0.04
BMI z-score	0.42 (0.87)	0.01	0.07	-0.16*	-0.02
FMI z-score [d]	-0.05 (0.95)	0.04	0.03	-0.09	-0.02

*Significant ($p < 0.05$). Abbreviations: NSP, non-starch polysaccharides; DXA, dual-energy x-ray absorptiometry; BMI z-score, body mass index z-score; FMI z-score, fat mass index z-score. [a] Correlation compares the variables with Shannon Index and gut microbiota profiles, determined using Spearman's correlations. [b] Describing microbial richness and evenness. [c] Missing data for n = 14. [d] Missing data for n = 38.

By contrast, profile 2 was associated with lower child BMI z-score (**Table 7.2**). Profile 3 was associated with the intakes of several nutrients (protein, dietary fibre, total NSP, insoluble NSP) and food groups ('nuts, seeds, legumes' and 'meat, fish, poultry') (**Table 7.2**).

When linear regression models were used to adjust for the four POI study groups and the demographic variables that had been found to be possible confounders (parity and mode of delivery for profile 2, and household deprivation for profile 3), statistically significant associations remained for: profile 2 and BMI z-score; and profile 3 and the dietary components dietary fibre, total NSP, and 'meat, fish, poultry' (**Table 7.3).**

In the fourth model (**Table 7.4**) the impact of including diet and BMI z-score variables in the same model was investigated. The association between profile 2 and BMI z-score was slightly strengthened by adjusting for some dietary components. For the dietary components tested in the regression model (dietary fibre, total NSP and 'meat, fish, poultry'), the biggest change in estimate, which was still small, was found for 'meat, fish, poultry' (estimate (95% CI): before adjustment -0.48 (-0.89, -0.07) after adjustment -0.52 (-0.94, -0.11)). The association between profile 3 and the dietary components fibre, total NSP and 'meat, fish, poultry' remained unchanged after adjusting for BMI z-score.

Alpha diversity described using the Shannon Index had a mean (SD) of 5.6 (0.5). None of the nutrient intake variables were associated with alpha diversity. Higher intakes of the food groups 'higher fibre more healthy cereals' and 'yoghurt', and lower intakes of 'lower fibre more healthy cereals', were correlated with greater alpha diversity (**Table 7.2**). However, following adjustment for the four POI study groups in the model, only the negative relationship with 'lower fibre more healthy cereals' remained statistically significant (**Table 7.3**).

Our data did not show a significant association between BMI z-score and the F/B ratio, either before (estimate (95% CI): 3.8 (-2.8, 10.5); $p = 0.260$) or after (estimate (95% CI): 2.7 (-4.0, 9.5); $p = 0.423$) adjustment for the four POI study groups.

Table 7.3: Associations of Shannon Index and gut microbiota profiles with selected variables with adjustment for covariates in the models.

Variables [a]	Shannon Index [b] Estimate (95% CI)	p-value	Profile 1 Estimate (95% CI)	p-value	Profile 2 Estimate (95% CI)	p-value	Profile 3 Estimate (95% CI)	p-value
First Model [c]								
Maternal parity								
Multi- vs primi-parous	0.00 (-0.11, 0.11)	0.960	0.38 (-0.33, 1.09)	0.297	-1.05 (-1.75, -0.35)	**0.003**	0.38 (-0.30, 1.05)	0.273
Household deprivation								
Low (1-3)	Reference		Reference		Reference		Reference	
Mid (4-7)	0.01 (-0.11, 0.13)	0.831	-0.06 (-0.85, 0.72)	0.876	0.06 (-0.73, 0.84)	0.885	-0.79 (-1.53, -0.06)	**0.035**
High (8-10)	0.04 (-0.11, 0.20)	0.581	-0.04 (-1.08, 0.99)	0.938	-0.06 (-1.09, 0.98)	0.916	-0.71 (-1.69, 0.26)	0.151
Mode of delivery								
Caesarean vs vaginal	-0.04 (-0.16, 0.08)	0.464	0.37 (-0.42, 1.16)	0.359	0.55 (-0.23, 1.34)	0.165	-0.19 (-0.93, 0.56)	0.622
Second Model [d]								
Any breastfeeding, *wk*	0.00 (-0.00, 0.00)	0.412	0.00 (-0.01, 0.01)	0.473	0.01 (0.00, 0.02)	**0.006**	0.00 (-0.01, 0.01)	0.806
Protein, *g*	0.00 (-0.00, 0.00)	0.231	0.00 (-0.02, 0.02)	0.987	0.00 (-0.02, 0.02)	0.907	0.02 (-0.00, 0.03)	0.079
Total dietary fibre, *g*	0.01 (-0.00, 0.02)	0.139	-0.03 (-0.09, 0.02)	0.256	-0.05 (-0.11, 0.01)	0.112	0.07 (0.01, 0.12)	**0.017**
Total NSP, *g*	0.01 (-0.00, 0.02)	0.184	-0.04 (-0.11, 0.03)	0.242	-0.06 (-0.13, 0.02)	0.113	0.08 (0.01, 0.15)	**0.019**
Insoluble NSP, *g*	0.01 (-0.01, 0.04)	0.204	-0.08 (-0.23, 0.07)	0.299	-0.12 (-0.27, 0.03)	0.103	0.13 (-0.02, 0.27)	0.082
Higher fibre more healthy cereals, *10g*	0.01 (-0.00, 0.01)	0.084	-0.02 (-0.07, 0.03)	0.348	-0.03 (-0.08, 0.02)	0.191	0.03 (-0.02, 0.08)	0.238
Lower fibre more healthy cereals, *10g*	-0.01 (-0.01, -0.00)	**0.028**	0.01 (-0.04, 0.05)	0.698	0.04 (-0.00, 0.09)	0.072	0.00 (-0.04, 0.05)	0.915
Nuts, seeds, legumes, *10g*	0.01 (-0.01, 0.03)	0.453	-0.08 (-0.22, 0.06)	0.242	-0.01 (-0.15, 0.13)	0.879	0.09 (-0.04, 0.22)	0.153
Fruits, *10g*	0.00 (-0.00, 0.01)	0.117	-0.02 (-0.05, 0.01)	0.121	-0.03 (-0.05, -0.00)	**0.048**	0.02 (-0.01, 0.04)	0.159
Potatoes, *10g*	-0.03 (-0.08, 0.01)	0.122	0.27 (-0.02, 0.55)	0.069	0.24 (-0.05, 0.53)	0.100	-0.11 (-0.38, 0.16)	0.427

Yoghurt, *10g*	0.01 (-0.00, 0.02)	0.138	-0.04 (-0.12, 0.03)	0.241	-0.02 (-0.09, 0.05)	0.603	0.02 (-0.05, 0.09)	0.613
Meat, fish, poultry, *10g*	0.00 (-0.01, 0.02)	0.598	0.03 (-0.06, 0.11)	0.555	0.05 (-0.04, 0.13)	0.283	0.10 (0.02, 0.18)	**0.019**
Third Model [e]								
BMI z-score	0.03 (-0.03, 0.09)	0.374	0.20 (-0.21, 0.62)	0.343	-0.48 (-0.89, -0.07)	**0.022**	0.06 (-0.33, 0.45)	0.759
FMI z-score [f]	0.04 (-0.02, 0.10)	0.232	0.18 (-0.23, 0.59)	0.387	-0.25 (-0.64, 0.15)	0.223	-0.04 (-0.42, 0.34)	0.833

Bold = significant ($p < 0.05$). Confounders (underlined) = $p < 0.25$. Abbreviations: vs, versus or compared to; NSP, non-starch polysaccharides; BMI z-score, body mass index z-score; FMI z-score, fat mass index z-score. [a] Variables were selected based on Spearman's correlations ≥ 0.09 from Table 2. [b] Describing microbial richness and evenness. [c] First model: Demographics variables (adjusted for four POI study groups). [d] Second model: Diet variables (adjusted for four POI study groups and other profile dependent demographic confounders). [e] Third model: BMI z-score and FMI z-score (adjusted for four POI study groups and other profile dependent demographic confounders). [f] Missing data for n = 38.

Table 7.4: Associations of gut microbiota profiles 2 and 3 after adjustments.

	Profile 2		Profile 3	
Variables	**Estimate (95% CI)**	**p-value**	**Estimate (95% CI)**	**p-value**
Fourth Model				
BMI z-score adjusted for total dietary fibre [a]	-0.45 (-0.76, -0.04)	**0.033**		
BMI z-score adjusted for total NSP [a]	-0.45 (-0.86, -0.04)	**0.033**		
BMI z-score adjusted for meat, fish, poultry [a]	-0.52 (-0.94, -0.11)	**0.013**		
Total dietary fibre adjusted for BMI z-score [b]			0.07 (0.01, 0.12)	**0.018**
Total NSP adjusted for BMI z-score [b]			0.08 (0.01, 0.15)	**0.021**
Meat, fish, poultry adjusted for BMI z-score [b]			0.10 (0.02, 0.18)	**0.020**

Bold = significant ($p < 0.05$). Abbreviations: NSP, non-starch polysaccharides; BMI z-score, body mass index z-score; FMI z-score, fat mass index z-score. [a] Also adjusted for: four POI study groups, parity, and mode of delivery. [b] Also adjusted for: four POI study groups, and household deprivation.

7.4 Discussion

This study identified three gut microbiota profiles in 5-year old New Zealand children, one associated with BMI z-score (profile 2) and one associated with several aspects of dietary intake, particularly those relating to dietary fibre (profile 3). The use of compositional PCA enabled the specific bacterial taxa characterizing each profile to be identified, and the strength of their contribution to the profile to be quantified.

In the rapidly growing area investigating associations between gut microbiota, diet, and health, this chapter of the thesis introduces a statistical approach that nutritionists are familiar with - using PCA to determine and describe patterns in complex dietary data. Compositional PCA has been used by some microbiologists (Gloor & Reid, 2016), but in this analysis, an approach to determining gut microbiota profiles is proposed that is analogous to the way dietary patterns are derived in nutrition studies while also accounting for the compositional nature of relative abundance data. Moreover, we have described the different profiles obtained and further analysed the associations of each profile with diet, body composition, and BMI z-score and FMI z-score in a way that does not appear to have been used previously in the literature.

There are three key advantages to this approach: i) the ability to reduce large amounts of compositional data into meaningful variables that can be described as communities of bacterial taxa, ii) the ability to describe the extent to which the gut bacterial community of every participant aligns with the gut bacterial profiles identified, and iii) the ability to investigate associations between gut microbiota profiles and other variables of interest, such as demographics, diet, and body composition.

The first key advantage is that compositional PCA extracts meaningful profiles from large amounts of compositional data. As mentioned in section 2.7.1, high-throughput DNA sequencing produces large datasets which are sparse and complex in their correlation structure, and need to be summarized (Xia et al., 2018). In addition, gut microbiota data such as relative abundance are different from dietary data in that they are compositional (Gloor et al., 2017; Gloor & Reid, 2016; Quinn et al., 2018), so that if the abundance of one bacterial taxon changes, this affects the relative abundance of other taxa. Data that exist in a closed space like this violate the assumptions of traditional statistical methods, including standard PCA and regression modelling. To be

able to use such methods, we must first transform the data using log-ratio transformations. Centred log-ratios are an appropriate transformation for PCA because they open up the variables from a closed space while also retaining a representation of each bacterial taxon. Once this transformation has taken place, PCA can be carried out as usual.

The second key advantage is that assigning every child a score for each of the gut microbiota profiles identified by compositional PCA enables us to consider every child's gut microbiota in all analyses. This method provides advantages over other methods such as enterotyping, in which each child is assigned to a single enterotype, meaning that relationships are limited to assessing differences between groups of people (as illustrated in **Figure 7.1**). By using the scores for each profile generated by compositional PCA, each individual in this study has three continuous variables with which to describe their microbiota. This enables the assessment of relationships with other variables, such as diet, body composition, and environmental factors, across all the participants. This increases statistical power compared to assigning participants to categories, and also means that these relationships are being examined across the same participants – rather than in different groups of people (Knights et al., 2014).

The third key advantage is that our statistical approach was able to demonstrate associations between two of our three gut microbiota profiles and BMI z-score and diet. Profile 2 was negatively associated with BMI z-score even after adjustment for dietary components, suggesting that higher BMI z-score were associated with less *Bacteroides* and more uncultured Christensenellaceae and Ruminococcaceae, independent of diet. While these bacteria have been reported in several studies investigating the relationships between weight status and the gut microbiota (Castaner et al., 2018), the directions of the associations vary. Our study supports earlier work in children (Hu et al., 2015; Jang, Choi, Kang, Park, & Lee, 2017; Riva et al., 2017), which showed BMI z-score being negatively associated with *Bacteroides*. The positive association we found with uncultured Christensenellaceae contrasts with another study in adults which found that *Christensenellaceae minuta*, a cultured member of Christensenellaceae, was associated with leanness (Goodrich et al., 2014). Further variation in findings is illustrated by a study in infants which found no significant associations with Christensenellaceae (Tun et al., 2018).

Profile 3 was positively associated with a number of dietary components (fibre, total NSP, and 'meat, fish, poultry' intake), independent of BMI z-score. Therefore, a higher intake of these dietary components is associated with more *Faecalibacterium*, *Eubacterium* and *Roseburia*, independent of body weight. These 3 genera are butyrate-producing bacteria (Louis & Flint, 2009), with butyrate being one of the 3 main short chain fatty acids produced by the gut microbiota. Butyrate from microbial sources is considered important for host colonic health as it acts as an energy source for the epithelial cells and has anti-inflammatory properties (Hamer et al., 2008). The positive association with dietary fibre and total NSP can probably be attributed to *Faecalibacterium*, *Eubacterium*, and *Roseburia* being genera that metabolize dietary plant polysaccharides (David et al., 2014; Zhang, Yang, Liang, Jiao, & Zhao, 2018).

There was a weak negative association between alpha diversity, as quantified by the Shannon Index, and intake of 'lower fibre more healthy cereals' (which consists of food items such as white bread, white rice, and cornflakes), after adjustment for the four POI study groups. Theoretically, the 'lower fibre more healthy cereals' food group has a much lower dietary fibre composition (1.6 g /100 g) compared to the 'higher fibre more healthy cereals' food group (7.0 g /100 g) Table 5.1. This effect was small (estimate; 95% CI: -0.01; -0.01, -0.00) and as food groups are defined differently in different studies, direct comparisons with the literature are difficult. However, assuming that a higher intake of 'lower fibre more healthy cereals' (associated in our data with lower alpha diversity) is associated with lower fibre intake in the diet overall, our finding agrees with another study where lower fibre intake has been associated with lower alpha diversity (Laursen et al., 2016) and also the study in **Chapter 4** of this thesis.

However, our study does not support the relationship between the commonly examined F/B ratio and BMI z-score. This reflects the findings of a recent meta-analysis which reported no significant difference in the F/B ratio between obese and normal weight individuals (Walters et al., 2014), but conflicts with some earlier work in children (Hou et al., 2017).

The strengths of this study include the large sample size, which enables the appropriate use of PCA accounting for the large number of variables (i.e. number of genera) per child participant. We also accounted for the nature of the gut microbiota

data by using compositional data analyses, and profiled and analysed the gut microbiota as a whole community. Furthermore, we used an FFQ that has been validated for the nutrients and food groups of interest to the gut microbiota and in the age group studied.

Also noteworthy is that the mean daily intakes of nutrients were similar to those of the 5-year old children from the EAT5 study (**Chapter 5**). However, intakes of the food groups were different, with the POI study children in this chapter having slightly lower mean intakes of 'potatoes' and 'meat, fish, egg', and higher intakes of all the other listed food groups in **Table 7.2**.

However, inferences from this study are limited by the non-representative nature of the sample (5-year old New Zealand children from higher socioeconomic families were over-represented so caution may be needed when generalizing to less advantaged families). In addition, a small number of potential confounders were investigated. Hence, we cannot rule out other potential confounding, including suppression effects, explaining our results. As we used 16S rRNA gene sequencing, reliable species identification and the functionality of the gut microbiota present was not determined. Moreover, the FFQ is also unable to assess intakes of dietary fibre fractions of interest such as resistant starch and oligosaccharides.

This study shows that the gut microbiota in 5-year old children can be described in terms of three different profiles, two of which are independently associated with diet and BMI z-score. However, the current data do not find evidence for associations between demographics, body composition, BMI z-score, FMI z-score or dietary components and profile 1. Hence, there are other factors in relation to profile 1 that we have not been able to identify. Additional, sufficiently powered studies are required using the same approach as described here to identify further links between the gut, diet and obesity, and for developing the next generation of research in which the impact on body composition of dietary interventions that modify the gut microbiota is determined. This study has, however, demonstrated the usefulness of compositional PCA in gut microbiota and diet research.

8 Conclusions and recommendations

8.1 Summary and conclusions

The overall aim of this thesis was to determine the effects of diet on the composition of the gut microbiota in infants and young children using appropriate nutritional, microbiological, and statistical methods. There is a critical window of rapid change in the gut microbiota that exists between the end of solely milk feeding (i.e. the introduction of solids) at around 4 - 6 months of age (Fallani et al., 2011), until around 3 years of age, when the gut microbiota is considered to have developed into a more adult-like composition (Laursen et al., 2017). However, few studies have been carried out to examine the effects of diet on the composition of the gut microbiota in infants during the complementary feeding period and young children (Laursen et al., 2017). Moreover, none have looked at the impact of 'Baby-Led Weaning' (BLW) (an increasingly popular alternative to traditional spoon-feeding) on the gut microbiota, and none have been able to look at children's diet using a dietary assessment tool that has been validated to measure nutrients and food groups that would be expected to influence the composition of the gut microbiota. In addition to addressing these gaps in the rapidly growing research area examining the diet and composition of the gut microbiota in children, this thesis contributes two statistical approaches that are known in the nutrition field but are largely novel in the gut microbiota field, and a validated dietary assessment tool.

In **Chapter 4**, the Candidate's analysis of data from the Baby-Led Introduction to SolidS (BLISS) randomised controlled trial (RCT) demonstrated the usefulness of mediation models for determining the pathway of relationships between dietary components and the composition of the gut microbiota. In the BLISS study, the infants who followed a modified version of BLW consumed a more adult type diet and had a faecal microbiota with less complex composition at 12 months than infants following traditional spoon-feeding. Through mediation modelling, the Candidate found that lower intakes of 'fruit, vegetables, nuts and legumes' and 'dietary fibre' were partially responsible (explained 29% and 25% of the relationship respectively) for this lower alpha diversity. However, the difference in alpha diversity between groups was modest and, at this stage, cannot be assumed to relate to changes in child development or health. Larger, longer-term studies are required before conclusions can be made about

the possible impact of these differences, or whether infant feeding guidelines should recommend that infants following a baby-led approach to infant feeding consume more fruit, vegetables, nuts and legumes or dietary fibre than is currently the case. Mediation models are rarely used in gut microbiota studies (Paolella & Vajro, 2018; Xia et al., 2018). The use of mediation analysis is an important contribution as this method is able to test which sequence of effects leads to an outcome, and allow causal inference (Fairchild & McDaniel, 2017) in gut microbiota studies.

In **Chapter 5**, the Candidate determined the validity and reproducibility of a dietary assessment tool (the Eating Assessment in Toddlers at 5 years food frequency questionnaire (EAT5 FFQ)), and therefore its appropriateness for use in large studies looking at the diet and composition of the gut microbiota in children. To date no studies have validated an FFQ specifically designed to measure intake of nutrients and food groups that are of relevance to the gut microbiota. This validation study showed that the EAT5 FFQ had acceptable validity for ranking intakes of energy, carbohydrate, dietary fibre, total non-starch polysaccharides (NSP), and insoluble NSP, when compared with the 3-day WDR, and very good reproducibility when measured over four weeks. The FFQ was also suitable for assessing mean absolute intake of carbohydrate, dietary fibre, and total NSP. The EAT5 FFQ is therefore an appropriate dietary assessment tool for investigating intake of nutrients and food groups that would be expected to influence the composition of the gut microbiota in studies of young children.

In **Chapter 7**, this validated dietary assessment tool was used to obtain dietary data in a large study looking at the effects of diet on the composition of the gut microbiota in young children. The results of the Candidate's analysis of data from the Prevention of Overweight in Infancy (POI) RCT showed the usefulness of compositional principal component analysis (PCA) for describing the composition of the gut microbiota, and for examining relationships between gut microbiota and dietary components and body size. The Candidate found that composition of the gut microbiota in 5-year old children could be described in terms of three different profiles. Profile 1 (positive loadings on *Blautia* and *Bifidobacterium*; negative loadings on *Bacteroides*) was not related to diet or body size; profile 2 (positive loadings on *Bacteroides;* negative loadings on uncultured Christensenellaceae and Ruminococcaceae) was associated with a lower body mass index (BMI) z-score; and profile 3 (positive loadings on *Faecalibacterium*, *Eubacterium* and *Roseburia*) was associated with higher

intakes of dietary fibre, total NSP, insoluble NSP, protein, meat, and nuts, seeds and legumes. The use of compositional PCA and additional regression analysis using the scores from the gut microbiota profiles as described in this thesis is useful as it takes into account the compositional nature, and sparsity (with many zeros) of gut microbiota data.

As no mechanistic study was carried out, the Candidate speculate that the three profiles exist partly due to cross-feeding or possibly competition for key dietary substrates such as dietary fibre and fibre fractions such as oligosaccharides. The rationale of this speculation is through considering the profiles' characterizing bacteria. Profile 3 has positive loadings of *Faecalibacterium*, *Eubacterium*, *Roseburia* and many members of the Lachnospiraceae (Firmicutes) family, which are known butyrate producers. Profile 2 has positive loadings of *Bacteroides* (a known propionate producer) and some negative loadings of members of the Firmicutes phylum. In a relatively opposite profile to profile 2, profile 1 has positive loadings of *Blautia* and *Bifidobacterium* (known acetate producers), members of the Firmicutes phylum and negative loadings of the members of the Bacteroidetes phylum (Louis, Hold, & Flint, 2014; Rowland et al., 2018).

The association between dietary fibre and profile 3 is relatively straight forward as dietary fibre acts as substrates for the butyrate-producing bacteria. Speculating on the possible mechanism for the associations between profile 2 and BMI z-score, as *Bacteroides* are known propionate producers, an increase in profile 2 (increase *Bacteroides* and associated increase in propionate (De Vadder et al., 2014)) is related to a decrease in BMI z-score, which could be due to propionate being known to inhibit acetate conversion into lipid in the liver and adipose tissue, contributing to decreased hepatic triglyceride content; and inhibiting food intake through secretion of satiety-regulating gastrointestinal hormones (Liou et al., 2013; Rios-Covian, Salazar, Gueimonde, & de los Reyes-Gavilan, 2017). Lastly, the Candidate speculate that a possible rationale for the lack of associations with any dietary or body composition factors with profile 1 may be due to the limitations that the study did not collect information on amount of breast milk, FOS or HMO. This speculation comes about from the positive loadings of *Blautia* and *Bifidobacterium* which are known utilizers of FOS and HMO for growth (Gotoh et al., 2018; Mao et al., 2018).

Additional studies are required using the same approaches as described in this thesis to identify further links between the gut, diet and obesity, and to develop the next generation of research in which the impact on body composition of dietary interventions that modify gut microbiota composition is determined.

8.2 Recommendations for future research

This thesis illustrates the urgent need for more studies investigating the effects of diet on the gut microbiota in infants and children.

- In future studies and analyses, where mediation analysis is possible, mediation models should be used to help detect causal inferences between predictors and the gut microbiota.
- Food composition databases that have more food items with complete fibre fractions need to be established and made freely available to researchers.
- When food composition information is available, studies should aim to determine intake of fibre fractions, not just total fibre. In addition, where studies have sufficient power, food groups should be separated into more groups related to fibre, for example the 'higher and lower fibre groups for the breads and cereals' in **Chapter 7** instead of just 'breads and cereals' in **Chapter 4** (and commonly used elsewhere in the literature). This would allow more specific interpretation of the data and enable researchers to generate more in-depth practical advice from their results. Standardisation of food groups between studies would be challenging, but would be very useful for comparing and combining data from different studies.
- FFQs are useful for gathering dietary data in large studies. In future studies, it is strongly recommended that if an FFQ is to be used, it must first be validated for the nutrients and food groups that will be studied so that the researchers can have confidence in the dietary data the FFQ generates.
- There are currently many bioinformatic software packages available that create visually striking illustrations and generate p-values. However, it is important that researchers are aware of the underlying statistical principles and assumptions being applied by these software programs. In turn, software programmers should strive to take into account the complexity and compositional nature of gut microbiota data when developing the software, for instance by including ways

that the data can be adjusted for confounders, and including ways to transform the data so that the complex and compositional nature of the gut microbiota can be considered.

- Future studies should use not only 16S rRNA methods, but other metagenomic, metatranscriptomic, metaproteomic and metabolomic methods, ideally within randomised controlled trials that are designed to test hypotheses based on ecological concepts such as keystone species, ecological resilience and disturbances in the community. This is so that the effects of specific aspects of the diet on the functions and activities of the gut microbiota, and their impacts on health, can be determined.

References

Adolfsson, O., Meydani, S. N., & Russell, R. M. (2004). Yogurt and gut function. *The American Journal of Clinical Nutrition, 80*(2), 245-256. doi:10.1093/ajcn/80.2.245

Aguirre, M., & Venema, K. (2015). The use of fecal samples for studying human obesity. *European Journal of Epidemiology, 30*(9), 1067-1069. doi:10.1007/s10654-015-0048-z

Aitchison, J., & Egozcue, J. J. (2005). Compositional data analysis: where are we and where should we be heading? *Mathematical Geology, 37*(7), 829-850. doi:10.1007/s11004-005-7383-7

Akagawa, S., Tsuji, S., Onuma, C., Akagawa, Y., Yamaguchi, T., Yamagishi, M., . . . Kaneko, K. (2019). Effect of delivery mode and nutrition on gut microbiota in neonates. *Annals of Nutrition and Metabolism, 74*(2), 132-139. doi:10.1159/000496427

Allaband, C., McDonald, D., Vázquez-Baeza, Y., Minich, J. J., Tripathi, A., Brenner, D. A., . . . Knight, R. (2019). Microbiome 101: studying, analyzing, and interpreting gut microbiome data for clinicians. *Clinical Gastroenterology and Hepatology, 17*(2), 218-230. doi:10.1016/j.cgh.2018.09.017

Amarri, S., Benatti, F., Callegari, M. L., Shahkhalili, Y., Chauffard, F., Rochat, F., . . . Morelli, L. (2006). Changes of gut microbiota and immune markers during the complementary feeding period in healthy breast-fed infants. *Journal of Pediatric Gastroenterology and Nutrition, 42*(5), 488-495. doi:10.1097/01.mpg.0000221907.14523.6d

Arimond, M., & Ruel, M. T. (2004). Dietary diversity is associated with child nutritional status: evidence from 11 demographic and health surveys. *The Journal of Nutrition, 134*(10), 2579-2585. doi:10.1093/jn/134.10.2579

Arrieta, M.-C., Stiemsma, L. T., Amenyogbe, N., Brown, E. M., & Finlay, B. (2014). The intestinal microbiome in early life: health and disease. *Frontiers in Immunology, 5*(427). doi:10.3389/fimmu.2014.00427

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., . . . Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature, 473*(7346), 174-180. doi:10.1038/nature09944

Atkinson, J., Salmond, P., & Crampton, C. (2014). *NZDep2013 Index of deprivation user's manual*. Wellington: University of Otago.

Avershina, E., Storrø, O., Øien, T., Johnsen, R., Pope, P., & Rudi, K. (2014). Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *FEMS Microbiology Ecology, 87*(1), 280-290. doi:10.1111/1574-6941.12223

Azad, M. B., Konya, T., Maughan, H., Guttman, D. S., Field, C. J., Chari, R. S., . . . Kozyrskyj, A. L. (2013). Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *CMAJ: Canadian Medical Association Journal, 185*(5), 385-394. doi:10.1503/cmaj.121189

Bäckhed, F. (2011). Programming of host metabolism by the gut microbiota. *Annals of Nutrition and Metabolism, 58*(Suppl 2), 44-52. doi:10.1159/000328042

Bäckhed, F., Fraser, Claire M., Ringel, Y., Sanders, Mary E., Sartor, R. B., Sherman, Philip M., . . . Finlay, B. B. (2012). Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications. *Cell Host & Microbe, 12*(5), 611-622. doi:10.1016/j.chom.2012.10.012

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., & Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science, 307*(5717), 1915-1920. doi:10.1126/science.1104816

Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., . . . Wang, J. (2015). Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host & Microbe, 17*(5), 690-703. doi:10.1016/j.chom.2015.04.004

Bahl, M. I., Bergström, A., & Licht, T. R. (2012). Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis. *FEMS Microbiology Letters, 329*(2), 193-197. doi:10.1111/j.1574-6968.2012.02523.x

Ballard, O., & Morrow, A. L. (2013). Human milk composition: nutrients and bioactive factors. *Pediatric Clinics of North America, 60*(1), 49-74. doi:10.1016/j.pcl.2012.10.002

Baothman, O. A., Zamzami, M. A., Taher, I., Abubaker, J., & Abu-Farha, M. (2016). The role of gut microbiota in the development of obesity and diabetes. *Lipids in Health and Disease, 15*(1), 1-8. doi:10.1186/s12944-016-0278-4

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research. Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182. doi:10.1037/0022-3514.51.6.1173

Baxter, N. T., Schmidt, A. W., Venkataraman, A., Kim, K. S., Waldron, C., & Schmidt, T. M. (2019). Dynamics of human gut microbiota and short-chain fatty acids in response to dietary interventions with three fermentable fibers. *mBio, 10*(1), e02566. doi:10.1128/mBio.02566-18

Beal, J. A. (2016). Baby-Led Weaning. *MCN: American Journal of Maternal Child Nursing, 41*(6). doi:10.1097/nmc.0000000000000283

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300.

Bergman, E. N. (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiological Reviews, 70*(2), 567-590. doi:10.1152/physrev.1990.70.2.567

Bergstrom, A., Skov, T. H., Bahl, M. I., Roager, H. M., Christensen, L. B., Ejlerskov, K. T., . . . Licht, T. R. (2014). Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of Danish infants. *Applied and Environmental Microbiology, 80*(9), 2889-2900. doi:10.1128/aem.00342-14

Bernal, M. J., Periago, M. J., Martínez, R., Ortuño, I., Sánchez-Solís, M., Ros, G., . . . Abellán, P. (2013). Effects of infant cereals with different carbohydrate profiles on colonic function—randomised and double-blind clinical trial in infants aged between 6 and 12 months—pilot study. *European Journal of Pediatrics, 172*(11), 1535-1542. doi:10.1007/s00431-013-2079-3

Bezirtzoglou, E., Tsiotsias, A., & Welling, G. W. (2011). Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe, 17*(6), 478-482. doi:10.1016/j.anaerobe.2011.03.009

Bircher, L., Schwab, C., Geirnaert, A., & Lacroix, C. (2018). Cryopreservation of artificial gut microbiota produced with in vitro fermentation technology. *Microbial Biotechnology, 11*(1), 163-175. doi:10.1111/1751-7915.12844

Bland, M. J., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*(2), 135-160. doi:10.1177/096228029900800204

Blum, R. E., Wei, E. K., Rockett, H. R., Langeliers, J. D., Leppert, J., Gardner, J. D., & Colditz, G. A. (1999). Validation of a food frequency questionnaire in Native American and Caucasian children 1 to 5 years of age. *Maternal and Child Health Journal, 3*(3), 167-172. doi:10.1023/A:1022350023163

Bode, L. (2012). Human milk oligosaccharides: Every baby needs a sugar mama. *Glycobiology, 22*(9), 1147-1162. doi:10.1093/glycob/cws074

Bode, L., & Jantscher-Krenn, E. (2012). Structure-Function Relationships of Human Milk Oligosaccharides. *Advances in Nutrition, 3*(3), 383S-391S. doi:10.3945/an.111.001404

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs, 27*(4), 325-349. doi:10.2307/1942268

Brown, A., & Lee, M. (2011). A descriptive study investigating the use and nature of baby-led weaning in a UK sample of mothers. *Maternal & Child Nutrition, 7*(1), 34-47. doi:10.1111/j.1740-8709.2010.00243.x

Brown, A., & Lee, M. (2013). An exploration of experiences of mothers following a baby-led weaning style: developmental readiness for complementary foods. *Maternal & Child Nutrition, 9*(2), 233-243. doi:10.1111/j.1740-8709.2011.00360.x

Butte, N. F., Garza, C., Smith, E. O., & Nichols, B. L. (1983). Evaluation of the deuterium dilution technique against the test-weighing procedure for the determination of breast milk intake. *The American Journal of Clinical Nutrition, 37*(6), 996-1003. doi:10.1093/ajcn/37.6.996

Butte, N. F., Lopez-Alarcon, M. G., & Garza, C. (2002). Nutrient adequacy of exclusive breastfeeding for the term infant during the first six months of life. *Expert Consultation on the Optimal Duration of Exclusive Breastfeeding 2001.* Retrieved from http://www.who.int/iris/handle/10665/42519

Cade, J., Thompson, R., Burley, V., & Warm, D. (2002). Development, validation and utilisation of food-frequency questionnaires – a review. *Public Health Nutrition, 5*(4), 567-587. doi:10.1079/PHN2001318

Calvert, C., Cade, J., Barrett, J. H., & Woodhouse, A. (1997). Using cross-check questions to address the problem of mis-reporting of specific food groups on Food Frequency Questionnaires. *European Journal of Clinical Nutrition, 51*(10), 708-712.

Cameron, S. L., Heath, A.-L., & Taylor, R. (2012a). How feasible is Baby-Led Weaning as an approach to infant feeding? A review of the evidence. *Nutrients, 4*(11), 1575-1609. doi:10.3390/nu4111575

Cameron, S. L., Heath, A.-L. M., & Taylor, R. W. (2012b). Healthcare professionals' and mothers' knowledge of, attitudes to and experiences with, Baby-Led Weaning: a content analysis study. *BMJ Open, 2*(6), e001542. doi:10.1136/bmjopen-2012-001542

Cameron, S. L., Taylor, R. W., & Heath, A.-L. M. (2015). Development and pilot testing of Baby-Led Introduction to SolidS - a version of Baby-Led Weaning modified to address concerns about iron deficiency, growth faltering and choking. *BMC Pediatrics, 15*, 99. doi:10.1186/s12887-015-0422-8

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Knight, R. (2010). QIIME allows analysis of high-throughput

community sequencing data. *Nature Methods, 7*(5), 335-336. doi:10.1038/nmeth.f.303
Castaner, O., Goday, A., Park, Y.-M., Lee, S.-H., Magkos, F., Shiow, S.-A. T. E., & Schröder, H. (2018). The gut microbiome profile in obesity: a systematic review. *International Journal of Endocrinology, 2018*, 4095789. doi:10.1155/2018/4095789
Centanni, M., Hutchison, J. C., Carnachan, S. M., Daines, A. M., Kelly, W. J., Tannock, G. W., & Sims, I. M. (2017). Differential growth of bowel commensal Bacteroides species on plant xylans of differing structural complexity. *Carbohydrate Polymers, 157*, 1374-1382. doi:10.1016/j.carbpol.2016.11.017
Chen, W., Zhang, C. K., Cheng, Y., Zhang, S., & Zhao, H. (2013). A comparison of methods for clustering 16S rRNA sequences into OTUs. *PLoS One, 8*(8), e70837. doi:10.1371/journal.pone.0070837
Christensen, L., Roager, H. M., Astrup, A., & Hjorth, M. F. (2018). Microbial enterotypes in personalized nutrition and obesity management. *The American Journal of Clinical Nutrition, 108*(4), 645-651. doi:10.1093/ajcn/nqy175
Clemente, J. C., Ursell, L. K., Parfrey, L. W., & Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. *Cell, 148*(6), 1258-1270. doi:10.1016/j.cell.2012.01.035
Clouston, A. (2014). *The role of commercial processed baby foods in the diets of New Zealand toddlers.* (Master of Dietetics), University of Otago, Dunedin, New Zealand.
Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., . . . Tiedje, J. M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research, 37*, D141-D145. doi:10.1093/nar/gkn879
Cook, A., Pryer, J., & Shetty, P. (2000). The problem of accuracy in dietary surveys. Analysis of the over 65 UK National Diet and Nutrition Survey. *Journal of Epidemiology and Community Health, 54*(8), 611-616. doi:10.1136/jech.54.8.611
Cooper, D. N., Martin, R. J., & Keim, N. L. (2015). Does whole grain consumption alter gut microbiota and satiety? *Healthcare, 3*(2), 364-392. doi:10.3390/healthcare3020364
Coppa, G. V., Gabrielli, O., Pierani, P., Catassi, C., Carlucci, A., & Giorgi, P. L. (1993). Changes in carbohydrate composition in human milk over 4 months of lactation. *Pediatrics, 91*(3), 637.
Costea, P. I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M. J., Bushman, F. D., . . . Bork, P. (2018). Enterotypes in the landscape of gut microbial community composition. *Nature Microbiology, 3*(1), 8-16. doi:10.1038/s41564-017-0072-8
Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I., & Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science, 326*(5960), 1694-1697. doi:10.1126/science.1177486
Cummings, J., & Mann, J. (2012). Carbohydrates. In J. Mann & A. S. Truswell (Eds.), *Essentials of Human Nutrition* (4th ed.). New York: Oxford University Press.
D'Andrea, E., Jenkins, K., Mathews, M., & Roebothan, B. (2016). Baby-led weaning: a preliminary investigation. *Canadian Journal of Dietetic Practice and Research. Revue Canadienne de la Pratique et de la Recherche en Dietetique, 77*(2), 72-77. doi:10.3148/cjdpr-2015-045
da Costa, T. H. M., Haisma, H., Wells, J. C. K., Mander, A. P., Whitehead, R. G., & Bluck, L. J. C. (2010). How much human milk do infants consume? Data from 12 countries using a standardized stable isotope methodology. *The Journal of Nutrition, 140*(12), 2227-2232. doi:10.3945/jn.110.123489

Daniels, L. (2017). *Impact of a baby-led approach to complementary feeding on iron and zinc intake and status: A randomised controlled trial.* (Doctor of Philosophy), University of Otago, Dunedin, New Zealand.

Daniels, L., Heath, A.-L. M., Williams, S. M., Cameron, S. L., Fleming, E. A., Taylor, B. J., . . . Taylor, R. W. (2015). Baby-Led Introduction to SolidS (BLISS) study: a randomised controlled trial of a baby-led approach to complementary feeding. *BMC Pediatrics, 15*, 179. doi:10.1186/s12887-015-0491-8

Daniels, L., Taylor, R. W., Williams, S. M., Gibson, R. S., Fleming, E. A., Wheeler, B. J., . . . Heath, A.-L. M. (2018). Impact of a modified version of baby-led weaning on iron intake and status: a randomised controlled trial. *BMJ Open, 8*(6), e019036. doi:10.1136/bmjopen-2017-019036

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., . . . Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature, 505*(7484), 559-563. doi:10.1038/nature12820

de Boer, R., Peters, R., Gierveld, S., Schuurman, T., Kooistra-Smid, M., & Savelkoul, P. (2010). Improved detection of microbial DNA after bead-beating before DNA isolation. *Journal of Microbiological Methods, 80*(2), 209-211. doi:10.1016/j.mimet.2009.11.009

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., . . . Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America, 107*(33), 14691-14696. doi:10.1073/pnas.1005963107

De Filippo, C., Di Paola, M., Ramazzotti, M., Albanese, D., Pieraccini, G., Banci, E., . . . Lionetti, P. (2017). Diet, environments, and gut microbiota. A preliminary investigation in children living in rural and urban Burkina Faso and Italy. *Frontiers in Microbiology, 8*(1979). doi:10.3389/fmicb.2017.01979

de la Cuesta-Zuluaga, J., Mueller, N. T., Álvarez-Quintero, R., Velásquez-Mejía, E. P., Sierra, J. A., Corrales-Agudelo, V., . . . Escobar, J. S. (2018). Higher fecal short-chain fatty acid levels are associated with gut microbiome dysbiosis, obesity, hypertension and cardiometabolic disease risk factors. *Nutrients, 11*(1), 51. doi:10.3390/nu11010051

de Onis, M., Onyango, A. W., Van den Broeck, J., Chumlea, C. W., & Martorell, R. (2004). Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. *Food and Nutrition Bulletin, 25*(1), S27-S36. doi:10.1177/15648265040251S104

De Vadder, F., Kovatcheva-Datchary, P., Goncalves, D., Vinera, J., Zitoun, C., Duchampt, A., . . . Mithieux, G. (2014). Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits. *Cell, 156*(1), 84-96. doi:10.1016/j.cell.2013.12.016

Delisle Nyström, C., Henriksson, H., Alexandrou, C., Bergström, A., Bonn, S., Bälter, K., & Löf, M. (2017). Validation of an online food frequency questionnaire against doubly labelled water and 24 h dietary recalls in pre-school children. *Nutrients, 9*(1), 66. doi:10.3390/nu9010066

den Besten, G., van Eunen, K., Groen, A. K., Venema, K., Reijngoud, D.-J., & Bakker, B. M. (2013). The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *Journal of Lipid Research, 54*(9), 2325-2340. doi:10.1194/jlr.R036012

Dennison, B. A., Jenkins, P. L., & Rockwell, H. L. (2000). Development and validation of an instrument to assess child dietary fat intake. *Preventive Medicine, 31*(3), 214-224. doi:10.1006/pmed.2000.0701

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., . . . Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology, 72*(7), 5069-5072. doi:10.1128/AEM.03006-05

Dewey, K. G., Heinig, M. J., Nommsen, L. A., & Lonnerdal, B. (1991a). Adequacy of energy intake amoung breast-fed infants in the DARLING study: relationships to growth velocity, morbidity, and activity levels: Davis Area Reasearch on Lactation, Infant Nutrition and Growth. *Journal of Pediatrics, 119*(4), 538-547. doi:10.1016/S0022-3476(05)82401-1

Dewey, K. G., Heinig, M. J., Nommsen, L. A., & Lonnerdal, B. (1991b). Maternal versus infant factors related to breast milk intake and residual milk volume: the DARLING study. *Pediatrics, 87*(6), 829-837.

Dhariwal, A., Chong, J., Habib, S., King, I. L., Agellon, L. B., & Xia, J. (2017). MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Research, 45*(W1), W180-W188. doi:10.1093/nar/gkx295

Diether, N. E., & Willing, B. P. (2019). Microbial fermentation of dietary protein: an important factor in diet-microbe-host interaction. *Microorganisms, 7*(1), 19. doi:10.3390/microorganisms7010019

Domel, S. B. (1997). Self-reports of diet: how children remember what they have eaten. *The American Journal of Clinical Nutrition, 65*(4), 1148S-1152S. doi:10.1093/ajcn/65.4.1148S

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America, 107*(26), 11971-11975. doi:10.1073/pnas.1002601107

Donaldson, G. P., Lee, S. M., & Mazmanian, S. K. (2016). Gut biogeography of the bacterial microbiota. *Nature Reviews: Microbiology, 14*(1), 20-32. doi:10.1038/nrmicro3552

Dunn, S., Datta, A., Kallis, S., Law, E., Myers, C. E., & Whelan, K. (2011). Validation of a food frequency questionnaire to measure intakes of inulin and oligofructose. *European Journal of Clinical Nutrition, 65*(3), 402-408. doi:10.1038/ejcn.2010.272

Dutman, A. E., Stafleu, A., Kruizinga, A., Brants, H. A. M., Westerterp, K. R., Kistemaker, C., . . . Goldbohm, R. A. (2011). Validation of an FFQ and options for data processing using the doubly labelled water method in children. *Public Health Nutrition, 14*(3), 410-417. doi:10.1017/S1368980010002119

Earl, C. C., Smith, M. T., Lease, R. A., & Bundy, B. C. (2018). Polyvinylsulfonic acid: a low-cost RNase inhibitor for enhanced RNA preservation and cell-free protein translation. *Bioengineered, 9*(1), 90-97. doi:10.1080/21655979.2017.1313648

Emmett, P., Symes, C., Braddon, F., & Heaton, K. (1992). Validation of a new questionnaire for assessing habitual intake of starch, non-starch polysaccharides, sugars and alcohol. *Journal of Human Nutrition and Dietetics, 5*(4), 245-253. doi:10.1111/j.1365-277X.1992.tb00159.x

Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention science: the official journal of the Society for Prevention Research, 10*(2), 87-99. doi:10.1007/s11121-008-0109-6

Fairchild, A. J., & McDaniel, H. L. (2017). Best (but oft-forgotten) practices: mediation analysis. *The American Journal of Clinical Nutrition, 105*(6), 1259-1271. doi:10.3945/ajcn.117.152546

Fallani, M., Amarri, S., Uusijarvi, A., Adam, R., Khanna, S., Aguilera, M., . . . Edwards, C. A. (2011). Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology, 157*, 1385-1392. doi:10.1099/mic.0.042143-0

Fangupo, L. J., Heath, A.-L. M., Williams, S. M., Erickson Williams, L. W., Morison, B. J., Fleming, E. A., . . . Taylor, R. W. (2016). A baby-led approach to eating solids and risk of choking. *Pediatrics, 138*(4), e20160772. doi:10.1542/peds.2016-0772

Fangupo, L. J., Heath, A. L. M., Williams, S. M., Somerville, M. R., Lawrence, J. A., Gray, A. R., . . . Taylor, R. W. (2015). Impact of an early-life intervention on the nutrition behaviors of 2-y-old children: A randomized controlled trial. *The American Journal of Clinical Nutrition, 102*(3), 704-712. doi:10.3945/ajcn.115.111823

Fernando, W., Hill, J., Zello, G., Tyler, R., Dahl, W., & Van Kessel, A. (2010). Diets supplemented with chickpea or its main oligosaccharide component raffinose modify faecal microbial composition in healthy adults. *Beneficial Microbes, 1*(2), 197-207. doi:10.3920/BM2009.0027

Ferrer, M., Méndez-García, C., Rojo, D., Barbas, C., & Moya, A. (2017). Antibiotic use and microbiome function. *Biochemical Pharmacology, 134*, 114-126. doi:10.1016/j.bcp.2016.09.007

Fisher, J. O., Butte, N. F., Mendoza, P. M., Wilson, T. A., Hodges, E. A., Reidy, K. C., & Deming, D. (2008). Overestimation of infant and toddler energy intake by 24-h recall compared with weighed food records. *The American Journal of Clinical Nutrition, 88*(2), 407-415. doi:10.1093/ajcn/88.2.407

Flint, H. J., Duncan, S. H., & Louis, P. (2017). The impact of nutrition on intestinal bacterial communities. *Current Opinion in Microbiology, 38*, 59-65. doi:10.1016/j.mib.2017.04.005

Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P., & Forano, E. (2012a). Microbial degradation of complex carbohydrates in the gut. *Gut Microbes, 3*(4), 289-306. doi:10.4161/gmic.19897

Flint, H. J., Scott, K. P., Louis, P., & Duncan, S. H. (2012b). The role of the gut microbiota in nutrition and health. *Nature Reviews: Gastroenterology & Hepatology, 9*(10), 577-589. doi:10.1038/nrgastro.2012.156

Flood, V. M., Wen, L. M., Hardy, L. L., Rissel, C., Simpson, J. M., & Baur, L. A. (2013). Reliability and validity of a short FFQ for assessing the dietary habits of 2–5-year-old children, Sydney, Australia. *Public Health Nutrition, 17*(3), 498-509. doi:10.1017/S1368980013000414

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286-299. doi:10.1037/1040-3590.7.3.286

Food Standards Australia New Zealand. (2008). *Addition of lutein as a nutritive substance to infant formula*. (Application A594). Food Standards Australia & New Zealand Retrieved from http://www.foodstandards.gov.au/code/infant/Pages/default.aspx

Food Standards Australia New Zealand. (2019). *Australian Food Composition Database – Release 1*. Canberra: FSANZ Retrieved from www.foodstandards.gov.au

Foster, E., Adamson, A. J., Anderson, A. S., Barton, K. L., & Wrieden, W. L. (2009). Estimation of portion size in children's dietary assessment: lessons learnt. *European Journal of Clinical Nutrition, 63*(S1), S45-S49. doi:10.1038/ejcn.2008.64

Fumagalli, F., Pontes Monteiro, J., Sartorelli, D. S., Vieira, M. N. C. M., & de Lourdes Pires Bianchi, M. (2008). Validation of a food frequency questionnaire for assessing dietary nutrients in Brazilian children 5 to 10 years of age. *Nutrition, 24*(5), 427-432. doi:10.1016/j.nut.2008.01.008

Garcia, A. L., McLean, K., & Wright, C. M. (2016). Types of fruits and vegetables used in commercial baby foods and their contribution to sugar content. *Maternal & Child Nutrition, 12*(4), 838-847. doi:10.1111/mcn.12208

Gibson, G. R., Scott, K. P., Rastall, R. A., Tuohy, K. M., Hotchkiss, A., Dubert-Ferrandon, A., . . . Buddington, R. (2010). Dietary prebiotics: current status and new definition. *Food Science and Technology Bulletin: Functional Foods, 7*(1), 1-19. doi:10.1616/1476-2137.15880

Gibson, R. S. (2005). *Principles of Nutritional Assessment* (2nd ed.). New York: Oxford University Press.

Gidrewicz, D. A., & Fenton, T. R. (2014). A systematic review and meta-analysis of the nutrient content of preterm and term breast milk. *BMC Pediatrics, 14*, 216. doi:10.1186/1471-2431-14-216

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: and this is not optional. *Frontiers in Microbiology, 8*(2224). doi:10.3389/fmicb.2017.02224

Gloor, G. B., & Reid, G. (2016). Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Canadian Journal of Microbiology, 62*(8), 692-703. doi:10.1139/cjm-2015-0821

Gondolf, U. H., Tetens, I., Hills, A. P., Michaelsen, K. F., & Trolle, E. (2012). Validation of a pre-coded food record for infants and young children. *European Journal of Clinical Nutrition, 66*(1), 91-96. doi:10.1038/ejcn.2011.133

Goodrich, J. K., Waters, Jillian L., Poole, Angela C., Sutter, Jessica L., Koren, O., Blekhman, R., . . . Ley, Ruth E. (2014). Human genetics shape the gut microbiome. *Cell, 159*(4), 789-799. doi:10.1016/j.cell.2014.09.053

Gorvitovskaia, A., Holmes, S. P., & Huse, S. M. (2016). Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle. *Microbiome, 4*, 15. doi:10.1186/s40168-016-0160-7

Gorzelak, M. A., Gill, S. K., Tasnim, N., Ahmadi-Vand, Z., Jay, M., & Gibson, D. L. (2015). Methods for improving human gut microbiome data by reducing variability through sample processing and storage of stool. *PLoS One, 10*(8), e0134802. doi:10.1371/journal.pone.0134802

Gotoh, A., Katoh, T., Sakanaka, M., Ling, Y., Yamada, C., Asakuma, S., . . . Katayama, T. (2018). Sharing of human milk oligosaccharides degradants within bifidobacterial communities in faecal cultures supplemented with Bifidobacterium bifidum. *Scientific Reports, 8*(1), 13958. doi:10.1038/s41598-018-32080-3

Gower, J. C. (2005). Principal coordinates analysis. In *Encyclopedia of Biostatistics*. Retrieved from https://doi.org/10.1002/0470011815.b2a13070

Graf, D., Di Cagno, R., Fak, F., Flint, H. J., Nyman, M., Saarela, M., & Watzl, B. (2015). Contribution of diet to the composition of the human gut microbiota. *Microbial Ecology in Health and Disease, 26*, 26164. doi:10.3402/mehd.v26.26164

Gratz, S. W., Wallace, R. J., & El-Nezami, H. S. (2011). Recent perspectives on the relations between fecal mutagenicity, genotoxicity, and diet. *Frontiers in Pharmacology, 2*(4). doi:10.3389/fphar.2011.00004

Guaraldi, F., & Salvatori, G. (2012). Effect of breast and formula feeding on gut microbiota shaping in newborns. *Frontiers in Cellular and Infection Microbiology, 2*, 94. doi:10.3389/fcimb.2012.00094

Hamaker, B. R., & Tuncil, Y. E. (2014). A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota. *Journal of Molecular Biology, 426*(23), 3838-3850. doi:10.1016/j.jmb.2014.07.028

Hamer, H. M., Jonkers, D., Venema, K., Vanhoutvin, S., Troost, F. J., & Brummer, R. J. (2008). Review article: the role of butyrate on colonic function. *Alimentary Pharmacology and Therapeutics, 27*(2), 104-119. doi:10.1111/j.1365-2036.2007.03562.x

Harttig, U., Haubrock, J., Knüppel, S., & Boeing, H. (2011). The MSM program: web-based statistics package for estimating usual dietary intake using the Multiple Source Method. *European Journal of Clinical Nutrition, 65*, S87-S91. doi:10.1038/ejcn.2011.92

Healey, G., Brough, L., Murphy, R., Hedderley, D., Butts, C., & Coad, J. (2016). Validity and reproducibility of a habitual dietary fibre intake short food frequency questionnaire. *Nutrients, 8*, 558. doi:10.3390/nu8090558

Heath, A.-L. M., Skeaff, C. M., & Gibson, R. S. (1999). Validation of a questionnaire method for estimating extent of menstrual blood loss in young adult women. *Journal of Trace Elements in Medicine and Biology, 12*(4), 231-235. doi:10.1016/S0946-672X(99)80063-7

Heiman, M. L., & Greenway, F. L. (2016). A healthy gastrointestinal microbiome is dependent on dietary diversity. *Molecular metabolism, 5*(5), 317-320. doi:10.1016/j.molmet.2016.02.005

Helnæs, A., Kyrø, C., Andersen, I., Lacoppidan, S., Overvad, K., Christensen, J., . . . Olsen, A. (2016). Intake of whole grains is associated with lower risk of myocardial infarction: the Danish Diet, Cancer and Health Cohort. *The American Journal of Clinical Nutrition, 103*(4), 999-1007. doi:10.3945/ajcn.115.124271

Ho, N. T., Li, F., Lee-Sarwar, K. A., Tun, H. M., Brown, B. P., Pannaraj, P. S., . . . Kuhn, L. (2018). Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. *Nature Communications, 9*(1), 4169-4169. doi:10.1038/s41467-018-06473-x

Hollister, E. B., Riehle, K., Luna, R. A., Weidler, E. M., Rubio-Gonzales, M., Mistretta, T.-A., . . . Versalovic, J. (2015). Structure and function of the healthy pre-adolescent pediatric gut microbiome. *Microbiome, 3*, 36. doi:10.1186/s40168-015-0101-x

Holscher, H. D. (2017). Dietary fiber and prebiotics and the gastrointestinal microbiota. *Gut Microbes, 8*(2), 172-184. doi:10.1080/19490976.2017.1290756

Honda, K., & Littman, D. R. (2012). The microbiome in infectious disease and inflammation. *Annual Review of Immunology, 30*, 759-795. doi:10.1146/annurev-immunol-020711-074937

Horwitz, W., & Latimer, G. W. (2006). *Official methods of analysis of AOAC international* (18th ed.). Gaithersburg, Maryland: AOAC International.

Hou, Y.-P., He, Q.-Q., Ouyang, H.-M., Peng, H.-S., Wang, Q., Li, J., . . . Yin, A.-H. (2017). Human gut microbiota associated with obesity in Chinese children and adolescents. *BioMed Research International, 2017*, 7585989. doi:10.1155/2017/7585989

Hu, F. B., Rimm, E., Smith-Warner, S. A., Feskanich, D., Stampfer, M. J., Ascherio, A., . . . Willett, W. C. (1999). Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *The American Journal of Clinical Nutrition, 69*(2), 243-249. doi:10.1093/ajcn/69.2.243

Hu, H.-J., Park, S.-G., Jang, H. B., Choi, M.-G., Park, K.-H., Kang, J. H., . . . Cho, S.-H. (2015). Obesity alters the microbial community profile in Korean adolescents. *PLoS One, 10*(7), e0134333. doi:10.1371/journal.pone.0134333

Huber, P. J. (1981). *Robust Statistics*. New York: John Wiley & Sons.

Huybrechts, I., De Backer, G., De Bacquer, D., Maes, L., & De Henauw, S. (2009). Relative validity and reproducibility of a food-frequency questionnaire for estimating food intakes among Flemish preschoolers. *International Journal of Environmental Research and Public Health, 6*(1), 382-399. doi:10.3390/ijerph6010382

Huybrechts, I., De Bacquer, D., Matthys, C., De Backer, G., & De Henauw, S. (2006). Validity and reproducibility of a semi-quantitative food-frequency questionnaire for estimating calcium intake in Belgian preschool children. *British Journal of Nutrition, 95*(4), 802-816. doi:10.1079/BJN20051671

International Atomic Energy Agency. (2010). *Stable isotope technique to assess intake of human milk in breastfed infants* (Vol. 7). Vienna: International Atomic Energy Agency.

Jaeggi, T., Kortman, G. A., Moretti, D., Chassard, C., Holding, P., Dostal, A., . . . Zimmermann, M. B. (2015). Iron fortification adversely affects the gut microbiome, increases pathogen abundance and induces intestinal inflammation in Kenyan infants. *Gut, 64*(5), 731-742. doi:10.1136/gutjnl-2014-307720

Jakobsson, H. E., Abrahamsson, T. R., Jenmalm, M. C., Harris, K., Quince, C., Jernberg, C., . . . Andersson, A. F. (2014). Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut, 63*(4), 559-566. doi:10.1136/gutjnl-2012-303249

Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinical Microbiology, 45*(9), 2761-2764. doi:10.1128/JCM.01228-07

Jang, H. B., Choi, M.-K., Kang, J. H., Park, S. I., & Lee, H.-J. (2017). Association of dietary patterns with the fecal microbiota in Korean adolescents. *BMC Nutrition, 3*, 20. doi:10.1186/s40795-016-0125-z

Jeffery, I. B., Claesson, M. J., O'Toole, P. W., & Shanahan, F. (2012). Categorization of the gut microbiota: enterotypes or gradients? *Nature Reviews: Microbiology, 10*, 591-592. doi:10.1038/nrmicro2859

Johnstone, I. M., & Titterington, D. M. (2009). Statistical challenges of high-dimensional data. *Philosophical Transactions. Series A: Mathematical, Physical, and Engineering Sciences, 367*(1906), 4237-4253. doi:10.1098/rsta.2009.0159

Keim, N. L., & Martin, R. J. (2014). Dietary whole grain-microbiota interactions: insights into mechanisms for human health. *Advances in Nutrition, 5*(5), 556-557. doi:10.3945/an.114.006536

Kent, J. C., Mitoulas, L. R., Cregan, M. D., Ramsay, D. T., Doherty, D. A., & Hartmann, P. E. (2006). Volume and frequency of breastfeedings and fat content

of breast milk throughout the day. *Pediatrics, 117*(3), e387-e395. doi:10.1542/peds.2005-1417

Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., . . . Dorrestein, P. C. (2018). Best practices for analysing microbiomes. *Nature Reviews: Microbiology, 16*(7), 410-422. doi:10.1038/s41579-018-0029-9

Knights, D., Ward, T. L., McKinlay, C. E., Miller, H., Gonzalez, A., McDonald, D., & Knight, R. (2014). Rethinking "enterotypes". *Cell Host & Microbe, 16*(4), 433-437. doi:10.1016/j.chom.2014.09.013

Kobayashi, T., Kamimura, M., Imai, S., Toji, C., Okamoto, N., Fukui, M., & Date, C. (2011). Reproducibility and validity of the food frequency questionnaire for estimating habitual dietary intake in children and adolescents. *Nutrition Journal, 10*, 27. doi:10.1186/1475-2891-10-27

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., . . . Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America, 108*(suppl 1), 4578-4585. doi:10.1073/pnas.1000081107

Koh, A., De Vadder, F., Kovatcheva-Datchary, P., & Bäckhed, F. (2016). From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell, 165*(6), 1332-1345. doi:10.1016/j.cell.2016.05.041

Konstantinidis, K. T., & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America, 102*(7), 2567-2572. doi:10.1073/pnas.0409727102

Krebs, N. F., Sherlock, L. G., Westcott, J., Culbertson, D., Hambidge, K. M., Feazel, L. M., . . . Frank, D. N. (2013). Effects of different complementary feeding regimens on iron status and enteric microbiota in breastfed infants. *Journal of Pediatrics, 163*(2), 416-423. doi:10.1016/j.jpeds.2013.01.024

Kuczynski, J., Lauber, C. L., Walters, W. A., Parfrey, L. W., Clemente, J. C., Gevers, D., & Knight, R. (2012). Experimental and analytical tools for studying the human microbiome. *Nature Reviews: Genetics, 13*(1), 47-58. doi:10.1038/nrg3129

Kuczynski, J., Stombaugh, J., Walters, W. A., González, A., Caporaso, J. G., & Knight, R. (2011). Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Current protocols in bioinformatics, Unit 10.7*. doi:10.1002/0471250953.bi1007s36

La-Ongkham, O., Nakphaichit, M., Leelavatcharamas, V., Keawsompong, S., & Nitisinprasert, S. (2015). Distinct gut microbiota of healthy children from two different geographic regions of Thailand. *Archives of Microbiology, 197*(4), 561-573. doi:10.1007/s00203-015-1089-0

Lagier, J.-C., Edouard, S., Pagnier, I., Mediannikov, O., Drancourt, M., & Raoult, D. (2015a). Current and past strategies for bacterial culture in clinical microbiology. *Clinical Microbiology Reviews, 28*(1), 208-236. doi:10.1128/CMR.00110-14

Lagier, J.-C., Hugon, P., Khelaifia, S., Fournier, P.-E., La Scola, B., & Raoult, D. (2015b). The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clinical Microbiology Reviews, 28*(1), 237-264. doi:10.1128/CMR.00014-14

Lamuel-Raventos, R. M., & Onge, M.-P. S. (2017). Prebiotic nut compounds and human microbiota. *Critical Reviews in Food Science and Nutrition, 57*(14), 3154-3163. doi:10.1080/10408398.2015.1096763

Landon, S., Colyer, C., & Salman, H. (2012). *The Resistant Starch Report*. Retrieved from Food Australia Supplement: Goodman Fielder Ltd and National Starch Ltd

Langdon, A., Crook, N., & Dantas, G. (2016). The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. *Genome Medicine, 8*, 39. doi:10.1186/s13073-016-0294-z

Lattimer, J. M., & Haub, M. D. (2010). Effects of dietary fiber and its components on metabolic health. *Nutrients, 2*(12), 1266-1289. doi:10.3390/nu2121266

Laursen, M. F., Andersen, L. B. B., Michaelsen, K. F., Mølgaard, C., Trolle, E., Bahl, M. I., & Licht, T. R. (2016). Infant gut microbiota development is driven by transition to family foods independent of maternal obesity. *mSphere, 1*(1), e00069. doi:10.1128/mSphere.00069-15

Laursen, M. F., Bahl, M. I., Michaelsen, K. F., & Licht, T. R. (2017). First foods and gut microbes. *Frontiers in Microbiology, 8*, 356. doi:10.3389/fmicb.2017.00356

Laursen, M. F., Zachariassen, G., Bahl, M. I., Bergström, A., Høst, A., Michaelsen, K. F., & Licht, T. R. (2015). Having older siblings is associated with gut microbiota development during early childhood. *BMC Microbiology, 15*, 154. doi:10.1186/s12866-015-0477-6

Lawley, B., & Tannock, G. W. (2012). Nucleic acid-based methods to assess the composition and function of the bowel microbiota. *Gastroenterology Clinics of North America, 41*(4), 855-868. doi:10.1016/j.gtc.2012.08.010

Lay, C. (2009). Molecular tools for investigating the gut microbiota. In D. Charalampopoulos & R. A. Rastall (Eds.), *Prebiotics and Probiotics Science and Technology* (pp. 31-78). New York: Springer.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., . . . Pedersen, O. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature, 500*, 541. doi:10.1038/nature12506

Leong, C., Haszard, J. J., Lawley, B., Otal, A., Taylor, R. W., Szymlek-Gay, E. A., . . . Heath, A.-L. M. (2018a). Mediation analysis as a means of identifying dietary components that affect the fecal microbiota of infants weaned by modified baby-led, compared to traditional, approaches. *Applied and Environmental Microbiology, 84*(18), e00914-00918. doi:10.1128/aem.00914-18

Leong, C., Taylor, R., Haszard, J., Fleming, E., Tannock, G., Szymlek-Gay, E., . . . Heath, A.-L. (2018b). Relative validity and reproducibility of a food frequency questionnaire to assess nutrients and food groups of relevance to the gut microbiota in young children. *Nutrients, 10*(11), 1627. doi:10.3390/nu10111627

Levin, A. M., Sitarik, A. R., Havstad, S. L., Fujimura, K. E., Wegienka, G., Cassidy-Bushrow, A. E., . . . Johnson, C. C. (2016). Joint effects of pregnancy, sociocultural, and environmental factors on early life gut microbiome structure and diversity. *Scientific Reports, 6*, 31775. doi:10.1038/srep31775

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Human gut microbes associated with obesity. *Nature, 444*, 1022. doi:10.1038/4441022a

Li, H. (2015). Microbiome, metagenomics, and high-dimensional compositional data analysis. *Annual Review of Statistics and Its Application, 2*(1), 73-94. doi:10.1146/annurev-statistics-010814-020351

Li, J., Riaz Rajoka, M. S., Shao, D., Jiang, C., Jin, M., Huang, Q., . . . Shi, J. (2017). Strategies to increase the efficacy of using gut microbiota for the modulation of obesity. *Obesity Reviews, 18*(11), 1260-1271. doi:10.1111/obr.12590

Lightowler, H. J., & Davies, G. J. (2002). Assessment of iodine intake in vegans: weighed dietary record vs duplicate portion technique. *European Journal of Clinical Nutrition, 56*(8), 765-770. doi:10.1038/sj.ejcn.1601392

Liljeberg Elmståhl, H. (2002). Resistant starch content in a selection of starchy foods on the Swedish market. *European Journal of Clinical Nutrition, 56*(6), 500-505. doi:10.1038/sj.ejcn.1601338

Lin, R., Liu, W., Piao, M., & Zhu, H. (2017). A review of the relationship between the gut microbiota and amino acid metabolism. *Amino Acids, 49*(12), 2083-2090. doi:10.1007/s00726-017-2493-3

Liou, A. P., Paziuk, M., Luevano, J.-M., Machineni, S., Turnbaugh, P. J., & Kaplan, L. M. (2013). Conserved shifts in the gut microbiota due to gastric bypass reduce host weight and adiposity. *Science Translational Medicine, 5*(178). doi:10.1126/scitranslmed.3005687

Liu, Y., Qin, S., Song, Y., Feng, Y., Lv, N., Xue, Y., . . . Yang, H. (2019a). The perturbation of infant gut microbiota caused by cesarean delivery Is partially restored by exclusive breastfeeding. *Frontiers in Microbiology, 10*(598). doi:10.3389/fmicb.2019.00598

Liu, Z., Diana, A., Slater, C., Preston, T., Gibson, R. S., Houghton, L., & Duffull, S. B. (2019b). Development of a nonlinear hierarchical model to describe the disposition of deuterium in mother-infant pairs to assess exclusive breastfeeding practice. *Journal of Pharmacokinetics and Pharmacodynamics, 46*(1), 1-13. doi:10.1007/s10928-018-9613-x

Livingstone, M. B. E., & Black, A. E. (2003). Markers of the validity of reported energy intake. *The Journal of Nutrition, 133*(3), 895S-920S. doi:10.1093/jn/133.3.895S

Lloyd-Price, J., Abu-Ali, G., & Huttenhower, C. (2016). The healthy human microbiome. *Genome Medicine, 8*, 51. doi:10.1186/s13073-016-0307-y

Lombard, M. J., Steyn, N. P., Charlton, K. E., & Senekal, M. (2015). Application and interpretation of multiple statistical tests to evaluate validity of dietary intake assessment methods. *Nutrition Journal, 14*, 40. doi:10.1186/s12937-015-0027-y

Louis, P., & Flint, H. J. (2009). Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiology Letters, 294*(1), 1-8. doi:10.1111/j.1574-6968.2009.01514.x

Louis, P., Flint, H. J., & Michel, C. (2016). How to manipulate the microbiota: prebiotics. In A. Schwiertz (Ed.), *Microbiota of the Human Body: Advances in Experimental Medicine and Biology* (pp. 119-142). Switzerland: Springer International Publishing.

Louis, P., Hold, G. L., & Flint, H. J. (2014). The gut microbiota, bacterial metabolites and colorectal cancer. *Nature Reviews Microbiology, 12*(661). doi:10.1038/nrmicro3344

Lozupone, C., & Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology, 71*(12), 8228-8235. doi:10.1128/aem.71.12.8228-8235.2005

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *The ISME Journal, 5*(2), 169-172. doi:10.1038/ismej.2010.133

Lozupone, C., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature, 489*(7415), 220-230. doi:10.1038/nature11550

Lozzo, P., & Sanguinetti, E. (2018). Early dietary patterns and microbiota development: still a way to go from descriptive interactions to health-relevant solutions. *Frontiers in Nutrition, 5*(5). doi:10.3389/fnut.2018.00005

Luque, V., Escribano, J., Mendez-Riera, G., Schiess, S., Koletzko, B., Verduci, E., . . . Closa-Monasterolo, R. (2013). Methodological approaches for dietary intake assessment in formula-fed infants. *Journal of Pediatric Gastroenterology and Nutrition, 56*(3), 320-327. doi:10.1097/MPG.0b013e3182779a60

Lyte, M., Chapel, A., Lyte, J. M., Ai, Y., Proctor, A., Jane, J. L., & Phillips, G. J. (2016). Resistant starch alters the microbiota-gut brain axis: Implications for dietary modulation of behavior. *PLoS One, 11*(1). doi:10.1371/journal.pone.0146406

Makki, K., Deehan, E. C., Walter, J., & Bäckhed, F. (2018). The impact of dietary fiber on gut microbiota in host health and disease. *Cell Host & Microbe, 23*(6), 705-715. doi:10.1016/j.chom.2018.05.012

Mao, B., Gu, J., Li, D., Cui, S., Zhao, J., Zhang, H., & Chen, W. (2018). Effects of different doses of fructooligosaccharides (FOS) on the composition of mice fecal microbiota, especially the bifidobacterium composition. *Nutrients, 10*(8), 1105. doi:10.3390/nu10081105

Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D. A., Hirschfield, G. M., Hold, G., . . . Hart, A. (2016). The gut microbiota and host health: a new clinical frontier. *Gut, 65*(2), 330-339. doi:10.1136/gutjnl-2015-309990

Marchesi, J. R., & Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome, 3*, 31. doi:10.1186/s40168-015-0094-5

Marcobal, A., Barboza, M., Sonnenburg, E. D., Pudlo, N., Martens, E. C., Desai, P., . . . Sonnenburg, J. L. (2011). Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host & Microbe, 10*(5), 507-514. doi:10.1016/j.chom.2011.10.007

Marlett, J. A., & Cheung, T.-F. (1997). Database and quick methods of assessing typical dietary fiber intakes using data for 228 commonly consumed foods. *Journal of the American Dietetic Association, 97*(10), 1139-1151. doi:10.1016/S0002-8223(97)00275-7

Marlett, J. A., & Vollendorf, N. W. (1993). Dietary fiber content and composition of vegetables determined by two methods of analysis. *Journal of Agricultural and Food Chemistry, 41*(10), 1608-1612. doi:10.1021/jf00034a016

Marlett, J. A., & Vollendorf, N. W. (1994a). Dietary fiber content and composition of different forms of fruits. *Food Chemistry, 51*(1), 39-44. doi:10.1016/0308-8146(94)90045-0

Marlett, J. A., & Vollendorf, N. W. (1994b). Dietary fiber content of cereal and grain products determined by enzymatic-chemical and enzymatic-gravimetric methods. *Journal of Food Composition and Analysis, 7*, 23-36. doi:10.1006/jfca.1994.1003

Marshall, T. A., Eichenberger Gilmore, J. M., Broffitt, B., Levy, S. M., & Stumbo, P. J. (2003). Relative validation of a beverage frequency questionnaire in children ages 6 months through 5 years using 3-day food and beverage diaries. *Journal of the American Dietetic Association, 103*(6), 714-720. doi:10.1053/jada.2003.50137

Martínez, I., Kim, J., Duffy, P. R., Schlegel, V. L., & Walter, J. (2010). Resistant starches types 2 and 4 have differential effects on the composition of the fecal microbiota in human subjects. *PLoS One, 5*(11), e15046. doi:10.1371/journal.pone.0015046

Martinez, I., Lattimer, J. M., Hubach, K. L., Case, J. A., Yang, J., Weber, C. G., . . . Walter, J. (2013). Gut microbiome composition is linked to whole grain-induced immunological improvements. *The ISME Journal, 7*(2), 269-280. doi:10.1038/ismej.2012.104

Masson, L. F., McNeill, G., Tomany, J. O., Simpson, J. A., Peace, H. S., Wei, L., . . . Bolton-Smith, C. (2003). Statistical approaches for assessing the relative validity of a food-frequency questionnaire: use of correlation coefficients and the kappa statistic. *Public Health Nutrition, 6*(3), 313-321. doi:10.1079/PHN2002429

Matamoros, S., Gras-Leguen, C., Le Vacon, F., Potel, G., & de La Cochetiere, M. F. (2013). Development of intestinal microbiota in infants and its impact on health. *Trends in Microbiology, 21*(4), 167-173. doi:10.1016/j.tim.2012.12.001

Matos, S. M. A., Prado, M. S., Santos, C. A. S. T., D'Innocenzo, S., Assis, A. M. O., Dourado, L. S., . . . Barreto, M. L. (2012). Validation of a food frequency questionnaire for children and adolescents aged 4 to 11 years living in Salvador, Bahia. *Nutricion Hospitalaria, 27*(4), 1114-1119. doi:10.3305/nh.2012.27.4.5883

Maukonen, J., Simões, C. D., & Saarela, M. (2012). The currently used commercial DNA-extraction methods give different results of clostridial and actinobacterial populations derived from human fecal samples. *FEMS Microbiology Ecology, 79*(3), 697-708. doi:10.1111/j.1574-6941.2011.01257.x

McPherson, J. D. (2014). A defining decade in DNA sequencing. *Nature Methods, 11*(10), 1003-1005. doi:10.1038/nmeth.3106

Meng, C., Zeleznik, O. A., Thallinger, G. G., Kuster, B., Gholami, A. M., & Culhane, A. C. (2016). Dimension reduction techniques for the integrative analysis of multi-omics data. *Briefings in Bioinformatics, 17*(4), 628-641. doi:10.1093/bib/bbv108

Menni, C., Jackson, M. A., Pallister, T., Steves, C. J., Spector, T. D., & Valdes, A. M. (2017). Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. *International Journal of Obesity, 41*(7), 1099-1105. doi:10.1038/ijo.2017.66

Mickey, R. M., & Greenland, S. (1989). The impact of confounder selection criteria on effect estimation. *American Journal of Epidemiology, 129*(1), 125-137. doi:10.1093/oxfordjournals.aje.a115101

Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J., . . . Ventura, M. (2017). The first microbial colonizers of the human gut: composition, activities, and health implications of the infant gut microbiota. *Microbiology and Molecular Biology Reviews, 81*(4), e00036. doi:10.1128/MMBR.00036-17

Mills, V. C., Skidmore, P. M. L., Watson, E. O., Taylor, R. W., Fleming, E. A., & Heath, A.-L. M. (2015). Relative validity and reproducibility of a food frequency questionnaire for identifying the dietary patterns of toddlers in New Zealand. *Journal of the Academy of Nutrition and Dietetics, 115*(4), 551-558. doi:10.1016/j.jand.2014.09.016

Ministry of Health. (2003). *NZ Food NZ Children: Key results of the 2002 National Children's Nutrition Survey*. Wellington: Ministry of Health.

Ministry of Health. (2008). *Food and Nutrition Guidelines for Healthy Infants and Toddlers (Aged 0–2): A background paper*. Wellington: Ministry of Health

Ministry of Health. (2012). *Food and Nutrition Guidelines for Healthy Children and Young People (Aged 2–18 years): A background paper.* Retrieved from Wellington:

Ministry of Health. (2015). *Eating and Activity Guidelines for New Zealand Adults*. Wellington: Ministry of Health

Ministry of Health. (2017). *Well Child Tamariki Ora visits*. Wellington: Ministry of Health Retrieved from https://www.health.govt.nz/your-health/pregnancy-and-kids/services-and-support-you-and-your-child/well-child-tamariki-ora-visits

Ministry of Health. (2019). *Annual Data Explorer 2017/18: New Zealand Health Survey*. Wellington: Ministry of Health Retrieved from https://minhealthnz.shinyapps.io/nz-health-survey-2017-18-annual-data-explorer

Moghames, P., Hammami, N., Hwalla, N., Yazbeck, N., Shoaib, H., Nasreddine, L., & Naja, F. (2016). Validity and reliability of a food frequency questionnaire to estimate dietary intake among Lebanese children. *Nutrition Journal, 15*, 4. doi:10.1186/s12937-015-0121-1

Morison, B. (2015). *Does a 'baby led' approach to complementary feeding have an impact on intake of selected nutrients, food preference and food variety in 12 month old infants?* (Master of Science), University of Otago, Dunedin, New Zealand.

Morison, B., Taylor, R. W., Haszard, J. J., Schramm, C. J., Williams Erickson, L., Fangupo, L. J., . . . Heath, A.-L. M. (2016). How different are baby-led weaning and conventional complementary feeding? A cross-sectional study of infants aged 6–8 months. *BMJ Open, 6*(5), e010665. doi:10.1136/bmjopen-2015-010665

Murphy, E. F., Cotter, P. D., Healy, S., Marques, T. M., Sullivan, O., Fouhy, F., . . . Shanahan, F. (2010). Composition and energy harvesting capacity of the gut microbiota: relationship to diet, obesity and time in mouse models. *Gut, 59*(12), 1635-1642. doi:10.1136/gut.2010.215665

Murphy, M. M., Douglass, J. S., & Birkett, A. (2008). Resistant starch intakes in the United States. *Journal of the American Dietetic Association, 108*(1), 67-78. doi:10.1016/j.jada.2007.10.012

Nakayama, J., Watanabe, K., Jiang, J., Matsuda, K., Chao, S.-H., Haryono, P., . . . Lee, Y.-K. (2015). Diversity in gut bacterial community of school-age children in Asia. *Scientific Reports, 5*, 8397. doi:10.1038/srep08397

Nakayama, J., Yamamoto, A., Palermo-Conde, L. A., Higashi, K., Sonomoto, K., Tan, J., & Lee, Y.-K. (2017). Impact of westernized diet on gut microbiota in children on Leyte island. *Frontiers in Microbiology, 8*, 197. doi:10.3389/fmicb.2017.00197

New Zealand Institute for Plant & Food Research Limited, & Ministry of Health. (2010). New Zealand Food Composition Database: New Zealand FOODfiles™ 2010. Retrieved from http://www.foodcomposition.co.nz/foodfiles

New Zealand Institute for Plant & Food Research Limited, & Ministry of Health. (2014). New Zealand Food Composition Database: New Zealand FOODfiles™ 2014. Retrieved from http://www.foodcomposition.co.nz/foodfiles

Ni, J., Wu, G. D., Albenberg, L., & Tomov, V. T. (2017). Gut microbiota and IBD: causation or correlation? *Nature Reviews: Gastroenterology & Hepatology, 14*, 573. doi:10.1038/nrgastro.2017.88

Oldenburg, C. E., Sié, A., Coulibaly, B., Ouermi, L., Dah, C., Tapsoba, C., . . . Doan, T. (2018). Effect of commonly used pediatric antibiotics on gut microbial diversity in preschool children in Burkina Faso: a randomized clinical trial. *Open Forum Infectious Diseases, 5*(11). doi:10.1093/ofid/ofy289

Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R., & Stahl, D. A. (1986). Microbial ecology and evolution: a ribosomal RNA Approach. *Annual Review of Microbiology, 40*(1), 337-365. doi:10.1146/annurev.mi.40.100186.002005

Oono, R. (2017). A confidence interval analysis of sampling effort, sequencing depth, and taxonomic resolution of fungal community ecology in the era of high-throughput sequencing. *PLoS One, 12*(12), e0189796. doi:10.1371/journal.pone.0189796

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A., & Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS Biology, 5*(7), e177. doi:10.1371/journal.pbio.0050177

Paolella, G., & Vajro, P. (2018). Maternal microbiota, prepregnancy weight, and mode of delivery: intergenerational transmission of risk for childhood overweight and obesity. *JAMA Pediatrics, 172*(4), 320-322. doi:10.1001/jamapediatrics.2017.5686

Parks, D. H., Tyson, G. W., Hugenholtz, P., & Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics, 30*(21), 3123-3124. doi:10.1093/bioinformatics/btu494

Parmanand, B. A., Kellingray, L., Le Gall, G., Basit, A. W., Fairweather-Tait, S., & Narbad, A. (2019). A decrease in iron availability to human gut microbiome reduces the growth of potentially pathogenic gut bacteria; an in vitro colonic fermentation study. *The Journal of Nutritional Biochemistry, 67*, 20-27. doi:10.1016/j.jnutbio.2019.01.010

Parrish, L. A., Marshall, J. A., Krebs, N. F., Rewers, M., & Norris, J. M. (2003). Validation of a food frequency questionnaire in preschool children. *Epidemiology, 14*(2), 213-217. doi:10.1097/01.EDE.0000041256.12192.23

Penders, J., Gerhold, K., Thijs, C., Zimmermann, K., Wahn, U., Lau, S., & Hamelmann, E. (2014). New insights into the hygiene hypothesis in allergic diseases: Mediation of sibling and birth mode effects by the gut microbiota. *Gut Microbes, 5*(2), 239-244. doi:10.4161/gmic.27905

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., . . . Stobberingh, E. E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics, 118*(2), 511-521. doi:10.1542/peds.2005-2824

Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The madness of microbiome: attempting to find consensus "best practice" for 16S microbiome studies. *Applied and Environmental Microbiology, 84*(7), e02627-02617. doi:10.1128/AEM.02627-17

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research, 35*(21), 7188-7196. doi:10.1093/nar/gkm864

Public Health England. (2015). Composition of foods integrated dataset (CoFID). Retrieved from https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid

Qasem, W., Azad, M. B., Hossain, Z., Azad, E., Jorgensen, S., Castillo San Juan, S., . . . Friel, J. (2017). Assessment of complementary feeding of Canadian infants: effects on microbiome & oxidative stress, a randomized controlled trial. *BMC Pediatrics, 17*, 54. doi:10.1186/s12887-017-0805-0

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research, 41*, D590-596. doi:10.1093/nar/gks1219

Quinn, T. P., Erb, I., Richardson, M. F., & Crowley, T. M. (2018). Understanding sequencing data as compositions: an outlook and review. *Bioinformatics, 34*(16), 2870-2878. doi:10.1093/bioinformatics/bty175

Rainakari, A. I., Rita, H., Putkonen, T., & Pastell, H. (2016). New dietary fibre content results for cereals in the Nordic countries using AOAC 2011.25 method. *Journal of Food Composition and Analysis, 51*, 1-8. doi:10.1016/j.jfca.2016.06.001

Rapley, G. (2011). Baby-led weaning: transitioning to solid foods at the baby's own pace. *Community Practice, 84*(6), 20-23.

Rapley, G., & Murkett, T. (2008). *Baby-led weaning: helping your baby love good food*. London: Random House UK.

Reck, M., Tomasch, J., Deng, Z., Jarek, M., Husemann, P., Wagner-Döbler, I., & Consortium, C. (2015). Stool metatranscriptomics: A technical guideline for mRNA stabilisation and isolation. *BMC Genomics, 16*(1), 494-494. doi:10.1186/s12864-015-1694-y

Reeves, M. M., Winkler, E. A. H., & Eakin, E. G. (2015). Fat and fibre behaviour questionnaire: Reliability, relative validity and responsiveness to change in Australian adults with type 2 diabetes and/or hypertension. *Nutrition and Dietetics, 72*(4), 368-376. doi:10.1111/1747-0080.12160

Ringel, Y., & Carroll, I. M. (2009). Alterations in the intestinal microbiota and functional bowel symptoms. *Gastrointestinal Endoscopy Clinics of North America, 19*(1), 141-150. doi:10.1016/j.giec.2008.12.004

Ríos-Covián, D., Ruas-Madiedo, P., Margolles, A., Gueimonde, M., de los Reyes-Gavilán, C. G., & Salazar, N. (2016). Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. *Frontiers in Microbiology, 7*, 185. doi:10.3389/fmicb.2016.00185

Rios-Covian, D., Salazar, N., Gueimonde, M., & de los Reyes-Gavilan, C. G. (2017). Shaping the metabolism of intestinal Bacteroides population through diet to improve human health. *Frontiers in Microbiology, 8*, 376. doi:10.3389/fmicb.2017.00376

Riva, A., Borgo, F., Lassandro, C., Verduci, E., Morace, G., Borghi, E., & Berry, D. (2017). Pediatric obesity is associated with an altered gut microbiota and discordant shifts in Firmicutes populations. *Environmental Microbiology, 19*(1), 95-105. doi:10.1111/1462-2920.13463

Robles Alonso, V., & Guarner, F. (2013). Linking the gut microbiota to human health. *British Journal of Nutrition, 109*(suppl 2), S21-S26. doi:10.1017/s0007114512005235

Rodriguez, C. A., Smith, E. R., Villamor, E., Zavaleta, N., Respicio-Torres, G., Contreras, C., . . . Franke, M. F. (2017). Development and validation of a food frequency questionnaire to estimate intake among children and adolescents in Urban Peru. *Nutrients, 9*(10), e1121. doi:10.3390/nu9101121

Rodríguez, J. M., Murphy, K., Stanton, C., Ross, R. P., Kober, O. I., Juge, N., . . . Collado, M. C. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microbial Ecology in Health and Disease, 26*, 26050. doi:10.3402/mehd.v26.26050

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ, 4*, e2584. doi:10.7717/peerj.2584

Rowan, H., & Harris, C. (2012). Baby-led weaning and the family diet. A pilot study. *Appetite, 58*(3), 1046-1049. doi:10.1016/j.appet.2012.01.033

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., & Tuohy, K. (2018). Gut microbiota functions: metabolism of nutrients and other food components. *European Journal of Nutrition, 57*(1), 1-24. doi:10.1007/s00394-017-1445-8

Ruggles, K. V., Wang, J., Volkova, A., Contreras, M., Noya-Alarcon, O., Lander, O., . . . Dominguez-Bello, M. G. (2018). Changes in the gut microbiota of urban subjects during an immersion in the traditional diet and lifestyle of a rainforest village. *mSphere, 3*(4), e00193. doi:10.1128/mSphere.00193-18

Russell, W. R., Gratz, S. W., Duncan, S. H., Holtrop, G., Ince, J., Scobbie, L., . . . Flint, H. J. (2011). High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health. *The American Journal of Clinical Nutrition, 93*(5), 1062-1072. doi:10.3945/ajcn.110.002188

Rutayisire, E., Huang, K., Liu, Y., & Tao, F. (2016). The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review. *BMC Gastroenterology, 16*, 86. doi:10.1186/s12876-016-0498-0

Saeedi, P., Skeaff, S. A., Wong, J. E., & Skidmore, P. M. L. (2016). Reproducibility and relative validity of a short food frequency questionnaire in 9–10 year-old children. *Nutrients, 8*(5), 271. doi:10.3390/nu8050271

Sawicki, C. M., Livingston, K. A., Obin, M., Roberts, S. B., Chung, M., & McKeown, N. M. (2017). Dietary fiber and the human gut microbiota: Application of evidence mapping methodology. *Nutrients, 9*(2), 125. doi:10.3390/nu9020125

Schwiertz, A., Taras, D., Schäfer, K., Beijer, S., Bos, N. A., Donus, C., & Hardt, P. D. (2010). Microbiota and SCFA in lean and overweight healthy subjects. *Obesity, 18*(1), 190-195. doi:10.1038/oby.2009.167

Scott, J. A., Chih, T. Y., & Oddy, W. H. (2012). Food variety at 2 years of age is related to duration of breastfeeding. *Nutrients, 4*(10), 1464-1474. doi:10.3390/nu4101464

Scott, K. P., Martin, J. C., Duncan, S. H., & Flint, H. J. (2014). Prebiotic stimulation of human colonic butyrate-producing bacteria and bifidobacteria, in vitro. *FEMS Microbiology Ecology, 87*(1), 30-40. doi:10.1111/1574-6941.12186

Segata, N. (2015). Gut microbiome: westernization and the disappearance of intestinal diversity. *Current Biology, 25*(14), R611-R613. doi:10.1016/j.cub.2015.05.040

Sela, D. A., & Mills, D. A. (2010). Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. *Trends in Microbiology, 18*(7), 298-307. doi:10.1016/j.tim.2010.03.008

Serdula, M. K., Ivery, D., Coates, R. J., Freedman, D. S., Williamson, D. F., & Byers, T. (1993). Do obese children become obese adults? A review of the literature. *Preventive Medicine, 22*(2), 167-177. doi:10.1006/pmed.1993.1014

Shankar, J., Szpakowski, S., Solis, N. V., Mounaud, S., Liu, H., Losada, L., . . . Filler, S. G. (2015). A systematic evaluation of high-dimensional, ensemble-based regression for exploring large model spaces in microbiome analyses. *BMC Bioinformatics, 16*, 31. doi:10.1186/s12859-015-0467-6

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379-423. doi:10.1002/j.1538-7305.1948.tb01338.x

Shetty, S. A., Marathe, N. P., Lanjekar, V., Ranade, D., & Shouche, Y. S. (2013). Comparative genome analysis of megasphaera sp. reveals niche specialization and its potential role in the human gut. *PLoS One, 8*(11), e79353. doi:10.1371/journal.pone.0079353

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika, 73*(3), 751-754. doi:10.1093/biomet/73.3.751

Simonyté Sjödin, K., Domellöf, M., Lagerqvist, C., Hernell, O., Lönnerdal, B., Szymlek-Gay, E. A., . . . Lind, T. (2018). Administration of ferrous sulfate drops

has significant effects on the gut microbiota of iron-sufficient infants: a randomised controlled study. *Gut*. doi:10.1136/gutjnl-2018-316988
Simpson, E. H. (1949). Measurement of diversity. *Nature, 163*(4148), 688.
Singapore Health Promotion Board. (2010). *Report of the National Nutrition Survey 2010*. Singapore: Singapore Health Promotion Board
Singh, R. K., Chang, H.-W., Yan, D., Lee, K. M., Ucmak, D., Wong, K., . . . Liao, W. (2017). Influence of diet on the gut microbiome and implications for human health. *Journal of Translational Medicine, 15*, 73. doi:10.1186/s12967-017-1175-y
Singh, S. B., Madan, J., Coker, M., Hoen, A., Baker, E. R., Karagas, M. R., & Mueller, N. T. (2019). Does birth mode modify associations of maternal pre-pregnancy BMI and gestational weight gain with the infant gut microbiome? *International Journal of Obesity*. doi:10.1038/s41366-018-0273-0
Smith-Brown, P., Morrison, M., Krause, L., & Davies, P. S. W. (2016). Dairy and plant based food intakes are associated with altered faecal microbiota in 2 to 3 year old Australian children. *Scientific Reports, 6*, 32385. doi:10.1038/srep32385
So, D., Whelan, K., Rossi, M., Morrison, M., Holtmann, G., Kelly, J. T., . . . Campbell, K. L. (2018). Dietary fiber intervention on gut microbiota composition in healthy adults: a systematic review and meta-analysis. *The American Journal of Clinical Nutrition, 107*(6), 965-983. doi:10.1093/ajcn/nqy041
Stanislawski, M. A., Dabelea, D., Wagner, B. D., Iszatt, N., Dahl, C., Sontag, M. K., . . . Eggesbø, M. (2018). Gut microbiota in the first 2 years of life and the association with body mass index at age 12 in a Norwegian birth cohort. *mBio, 9*(5), e01751. doi:10.1128/mBio.01751-18
Stanislawski, M. A., Dabelea, D., Wagner, B. D., Sontag, M. K., Lozupone, C. A., & Eggesbø, M. (2017). Pre-pregnancy weight, gestational weight gain, and the gut microbiota of mothers and their infants. *Microbiome, 5*, 113. doi:10.1186/s40168-017-0332-0
Stein, A. D., Shea, S., Basch, C. E., Contento, I. R., & Zyberf, P. (1992). Consistency of the willett semiquantitative food frequency questionnaire and 24-hour dietary recalls in estimating nutrient intakes of preschool children. *American Journal of Epidemiology, 135*(6), 667-677. doi:10.1093/oxfordjournals.aje.a116346
Stinson, L. F., Payne, M. S., & Keelan, J. A. (2018). A critical review of the bacterial baptism hypothesis and the impact of cesarean delivery on the infant microbiome. *Frontiers in medicine, 5*, 135. doi:10.3389/fmed.2018.00135
Sugino, K. Y., Paneth, N., & Comstock, S. S. (2019). Michigan cohorts to determine associations of maternal pre-pregnancy body mass index with pregnancy and infant gastrointestinal microbial communities: Late pregnancy and early infancy. *PLoS One, 14*(3), e0213733. doi:10.1371/journal.pone.0213733
Szymlek-Gay, E. A., Ferguson, E. L., Heath, A.-L. M., & Fleming, E. A. (2010). Quantities of foods consumed by 12- to 24-month-old New Zealand children. *Nutrition & Dietetics, 67*(4), 244-250. doi:10.1111/j.1747-0080.2010.01471.x
Tanaka, M., & Nakayama, J. (2017). Development of the gut microbiota in infancy and its impact on health in later life. *Allergology International, 66*(4), 515-522. doi:10.1016/j.alit.2017.07.010
Tannock, G. W. (2005). *Probiotics and Prebiotics: Scientific Aspects*. Norfolk, UK: Caister Academic Press.
Tannock, G. W. (2017). *Understanding the Gut Microbiota*. Hoboken, NJ: John Wiley and Sons.

Tannock, G. W., Lawley, B., Munro, K., Gowri Pathmanathan, S., Zhou, S. J., Makrides, M., . . . Hodgkinson, A. J. (2013). Comparison of the compositions of the stool microbiotas of infants fed goat milk formula, cow milk-based formula, or breast milk. *Applied and Environmental Microbiology, 79*(9), 3040-3048. doi:10.1128/AEM.03910-12

Tap, J., Furet, J.-P., Bensaada, M., Philippe, C., Roth, H., Rabot, S., . . . Leclerc, M. (2015). Gut microbiota richness promotes its stability upon increased dietary fibre intake in healthy adults. *Environmental Microbiology, 17*(12), 4954-4964. doi:10.1111/1462-2920.13006

Taylor, B. J., Gray, A. R., Galland, B. C., Heath, A.-L. M., Lawrence, J., Sayers, R. M., . . . Taylor, R. W. (2017a). Targeting sleep, food, and activity in infants for obesity prevention: an RCT. *Pediatrics, 139*(3), e20162037. doi:10.1542/peds.2016-2037

Taylor, B. J., Heath, A.-L. M., Galland, B. C., Gray, A. R., Lawrence, J. A., Sayers, R. M., . . . Taylor, R. W. (2011). Prevention of Overweight in Infancy (POI.nz) study: a randomised controlled trial of sleep, food and activity interventions for preventing overweight from birth. *BMC Public Health, 11*, 942. doi:10.1186/1471-2458-11-942

Taylor, M. W. (2016). Devil in the detail: a closer look at childhood obesity and the gut microbiota. *Environmental Microbiology, 19*(1), 11-12. doi:10.1111/1462-2920.13540

Taylor, R. W., & Goulding, A. (1998). Validation of a short food frequency questionnaire to assess calcium intake in children aged 3 to 6 years. *European Journal of Clinical Nutrition, 52*, 464-465.

Taylor, R. W., Gray, A. R., Heath, A.-L. M., Galland, B. C., Lawrence, J., Sayers, R., . . . Taylor, B. J. (2018). Sleep, nutrition, and physical activity interventions to prevent obesity in infancy: follow-up of the Prevention of Overweight in Infancy (POI) randomized controlled trial at ages 3.5 and 5 y. *The American Journal of Clinical Nutrition, 108*(2), 228-236. doi:10.1093/ajcn/nqy090

Taylor, R. W., Heath, A.-L. M., Galland, B. C., Cameron, S. L., Lawrence, J. A., Gray, A. R., . . . Taylor, B. J. (2016). Three-year follow-up of a randomised controlled trial to reduce excessive weight gain in the first two years of life: protocol for the POI follow-up study. *BMC Public Health, 16*, 771. doi:10.1186/s12889-016-3383-4

Taylor, R. W., Williams, S. M., Fangupo, L. J., Wheeler, B. J., Taylor, B. J., Daniels, L., . . . Heath, A. M. (2017b). Effect of a baby-led approach to complementary feeding on infant growth and overweight: a randomized clinical trial. *JAMA Pediatrics, 171*(9), 1-9. doi:10.1001/jamapediatrics.2017.1284

The GBD Obesity Collaboration, Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., . . . Gakidou, E. (2014). Global, regional and national prevalence of overweight and obesity in children and adults 1980-2013: A systematic analysis. *Lancet, 384*(9945), 766-781. doi:10.1016/S0140-6736(14)60460-8

Theander, O., Åman, P., Westerlund, E., & Graham, H. (1990). The uppsala method for rapid analysis of total dietary fiber. In I. Furda & C. J. Brine (Eds.), *New Developments in Dietary Fiber. Advances in Experimental Medicine and Biology* (Vol. 270, pp. 273-281). Boston, MA: Springer.

Thompson, A. L., Monteagudo-Mera, A., Cadenas, M. B., Lampl, M. L., & Azcarate-Peril, M. A. (2015). Milk- and solid-feeding practices and daycare attendance are associated with differences in bacterial diversity, predominant communities, and

metabolic and immune function of the infant gut microbiome. *Frontiers in Cellular and Infection Microbiology, 5*(3). doi:10.3389/fcimb.2015.00003

Tsilimigras, M. C. B., & Fodor, A. A. (2016). Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Annals of Epidemiology, 26*(5), 330-335. doi:10.1016/j.annepidem.2016.03.002

Tun, H. M., Bridgman, S. L., Chari, R., Field, C. J., Guttman, D. S., Becker, A. B., . . . Investigators for the CHILD Study. (2018). Roles of birth mode and infant gut microbiota in intergenerational transmission of overweight and obesity from mother to offspring. *JAMA Pediatrics, 172*(4), 368-377. doi:10.1001/jamapediatrics.2017.5535

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature, 449*(7164), 804-810. doi:10.1038/nature06244

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature, 444*(7122), 1027-1131. doi:10.1038/nature05414

Turroni, F., Peano, C., Pass, D. A., Foroni, E., Severgnini, M., Claesson, M. J., . . . Ventura, M. (2012). Diversity of bifidobacteria within the infant gut microbiota. *PLoS One, 7*(5), e36957. doi:10.1371/journal.pone.0036957

University of Otago and Ministry of Health. (2011). Methodology Report for the 2008/09 NZ Adult Nutrition Survey. Retrieved from https://www.health.govt.nz/publication/methodology-report-2008-09-nz-adult-nutrition-survey

Ursell, L. K., Haiser, H. J., Van Treuren, W., Garg, N., Reddivari, L., Vanamala, J., . . . Knight, R. (2014). The intestinal metabolome: An intersection between microbiota and host. *Gastroenterology, 146*(6), 1470-1476. doi:10.1053/j.gastro.2014.03.001

USDA. (2017). USDA Food Composition Databases version 3.7. Retrieved from https://ndb.nal.usda.gov/ndb/

Valdes, A. M., Walter, J., Segal, E., & Spector, T. D. (2018). Role of the gut microbiota in nutrition and health. *British Medical Journal, 361*(suppl 1), 36-44. doi:10.1136/bmj.k2179

Vallès, Y., Artacho, A., Pascual-García, A., Ferrús, M. L., Gosalbes, M. J., Abellán, J. J., & Francino, M. P. (2014). Microbial succession in the gut: directional trends of taxonomic and functional change in a birth cohort of Spanish infants. *PLoS Genetics, 10*(6), e1004406. doi:10.1371/journal.pgen.1004406

Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A., & Knight, R. (2013). EMPeror: a tool for visualizing high-throughput microbial community data. *GigaScience, 2*, 16. doi:10.1186/2047-217x-2-16

Vereecken, C., Covents, M., & Maes, L. (2010). Comparison of a food frequency questionnaire with an online dietary assessment tool for assessing preschool children's dietary intake. *Journal of Human Nutrition and Dietetics, 23*(5), 502-510. doi:10.1111/j.1365-277X.2009.01038.x

Vioque, J., Gimenez-Monzo, D., Navarrete-Muñoz, E. M., Garcia-de-la-Hera, M., Gonzalez-Palacios, S., Rebagliato, M., . . . Members of NMA-Valencia Cohort Study. (2016). Reproducibility and validity of a food frequency questionnaire designed to assess diet in children aged 4-5 years. *PLoS One, 11*(11), e0167338. doi:10.1371/journal.pone.0167338

Vital, M., Howe, A., Bergeron, N., Krauss, R. M., Jansson, J. K., & Tiedje, J. M. (2018). Metagenomic insights into the degradation of resistant starch by human

gut microbiota. *Applied and Environmental Microbiology, 84*(23), e01562-01518. doi:10.1128/AEM.01562-18
Voreades, N., Kozil, A., & Weir, T. L. (2014). Diet and the development of the human intestinal microbiome. *Frontiers in Microbiology, 5*(494). doi:10.3389/fmicb.2014.00494
Walker, A. W. (2016). Studying the human microbiota. In A. Schwiertz (Ed.), *Microbiota of the Human Body. Advances in Experimental Medicine and Biology* (pp. 5-32). Switzerland: Springer International Publishing.
Walker, A. W., Martin, J. C., Scott, P., Parkhill, J., Flint, H. J., & Scott, K. P. (2015). 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome, 3*, 26. doi:10.1186/s40168-015-0087-4
Walters, W. A., Xu, Z., & Knight, R. (2014). Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Letters, 588*(22), 4223-4233. doi:10.1016/j.febslet.2014.09.039
Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician, 70*(2), 129-133. doi:10.1080/00031305.2016.1154108
Watson, E. O., Heath, A.-L. M., Taylor, R. W., Mills, V. C., Barris, A. C., & Skidmore, P. M. L. (2015). Relative validity and reproducibility of an FFQ to determine nutrient intakes of New Zealand toddlers aged 12–24 months. *Public Health Nutrition, 18*(18), 3265-3271. doi:10.1017/S1368980015000841
Watson, J. F., Collins, C. E., Sibbritt, D. W., Dibley, M. J., & Garg, M. L. (2009). Reproducibility and comparative validity of a food frequency questionnaire for Australian children and adolescents. *International Journal of Behavioral Nutrition and Physical Activity, 6*, 62. doi:10.1186/1479-5868-6-62
Weber, D. R., Moore, R. H., Leonard, M. B., & Zemel, B. S. (2013). Fat and lean BMI reference curves in children and adolescents and their utility in identifying excess adiposity compared with BMI and percentage body fat. *The American Journal of Clinical Nutrition, 98*(1), 49-56. doi:10.3945/ajcn.112.053611
Westenbrink, S., Brunt, K., & van der Kamp, J.-W. (2013). Dietary fibre: Challenges in production and use of food composition data. *Food Chemistry, 140*(3), 562-567. doi:10.1016/j.foodchem.2012.09.029
WHO. (2002). Complementary feeding. Report of the global consultation. Retrieved from http://www.who.int/nutrition/publications/Complementary_Feeding.pdf
WHO. (2019). Childhood overweight and obesity. *Global Strategy on Diet, Physical Activity and Health.* Retrieved from https://www.who.int/dietphysicalactivity/childhood/en/
WHO Multicentre Growth Reference Study Group. (2006). WHO child growth standards based on length/height, weight and age. *Acta Paediatrica, Suppl 450*, 76-85. doi:10.1080/08035320500495548
Willett, W. (1998). *Nutritional Epidemiology* (2nd ed.). New York: Oxford University Press.
Willey, J., Sherwood, L., & Woolverton, C. (2013). *Prescott's Microbiology* (9th ed.): McGraw-Hill Higher Education.
Williams, B. A., Grant, L. J., Gidley, M. J., & Mikkelsen, D. (2017). Gut fermentation of dietary fibres: physico-chemistry of plant cell walls and implications for health. *International Journal of Molecular Sciences, 18*(10), 2203. doi:10.3390/ijms18102203

Williams-Erickson, L. (2015). *A Baby-Led approach to complementary feeding: Adherence and infant food and nutrient intakes at seven months of age.* (Master of Science), University of Otago, Dunedin, New Zealand.
Williams-Erickson, L., Taylor, R., Haszard, J., Fleming, E., Daniels, L., Morison, B., . . . Heath, A.-L. M. (2018). Impact of a modified version of baby-led weaning on infant food and nutrient intakes: the BLISS randomized controlled trial. *Nutrients, 10*(6), 740. doi:10.3390/nu10060740
Wilmes, P., & Bond, P. L. (2009). Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the earth's biogeochemical cycles. *Current Opinion in Microbiology, 12*(3), 310-317. doi:10.1016/j.mib.2009.03.004
Woese, C. R. (1987). Bacterial evolution. *Microbiological Reviews, 51*(2), 221-271.
World Health Organization. (2019, 11 Febreuary 2019). Exclusive breastfeeding for optimal growth, development and health of infants. *e-Library of Evidence for Nutrition Actions (eLENA).* Retrieved from https://www.who.int/elena/titles/exclusive_breastfeeding/en/
Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y. Y., Keilbaugh, S. A., . . . Lewis, J. D. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science, 334*(6052), 105-108. doi:10.1126/science.1208344
Xia, Y., & Sun, J. (2017). Hypothesis testing and statistical analysis of microbiome. *Genes & Diseases, 4*(3), 138-148. doi:10.1016/j.gendis.2017.06.001
Xia, Y., Sun, J., & Chen, D. G. (2018). *Statistical Analysis of Microbiome Data with R* (J. Chen & D. G. Chen Eds.). Singapore: Springer Nature Singapore Pte Ltd.
Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., . . . Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature, 486*(7402), 222-227. doi:10.1038/nature11053
Zabel, B., Yde, C. C., Roos, P., Marcussen, J., Jensen, H. M., Salli, K., . . . Morovic, W. (2019). Novel genes and metabolite trends in Bifidobacterium longum subsp. infantis Bi-26 metabolism of human milk oligosaccharide 2′-fucosyllactose. *Scientific Reports, 9*(1), 7983. doi:10.1038/s41598-019-43780-9
Ze, X., Ben David, Y., Laverde-Gomez, J. A., Dassa, B., Sheridan, P. O., Duncan, S. H., . . . Flint, H. J. (2015). Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic Firmicutes bacterium Ruminococcus bromii. *mBio, 6*(5), e01058. doi:10.1128/mBio.01058-15
Ze, X., Duncan, S. H., Louis, P., & Flint, H. J. (2012). Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *The ISME Journal, 6*(8), 1535-1543. doi:10.1038/ismej.2012.4
Zhang, J., Wei, Z., & Chen, J. (2018). A distance-based approach for testing the mediation effect of the human microbiome. *Bioinformatics, 34*(11), 1875-1883. doi:10.1093/bioinformatics/bty014
Zhang, T., Yang, Y., Liang, Y., Jiao, X., & Zhao, C. (2018). Beneficial effect of intestinal fermentation of natural polysaccharides. *Nutrients, 10*(8), 1055. doi:10.3390/nu10081055
Zhang, X., Nieuwdorp, M., Groen, A. K., & Zwinderman, A. H. (2019). Statistical evaluation of diet-microbe associations. *BMC Microbiology, 19*, 90. doi:10.1186/s12866-019-1464-0
Zhong, H., Penders, J., Shi, Z., Ren, H., Cai, K., Fang, C., . . . Kristiansen, K. (2019). Impact of early events and lifestyle on the gut microbiota and metabolic phenotypes in young school-age children. *Microbiome, 7*, 2. doi:10.1186/s40168-018-0608-z

Zmora, N., Suez, J., & Elinav, E. (2019). You are what you eat: diet, health and the gut microbiota. *Nature Reviews: Gastroenterology & Hepatology, 16*, 35-56. doi:10.1038/s41575-018-0061-2

Zou, Y., Xue, W., Luo, G., Deng, Z., Qin, P., Guo, R., . . . Xiao, L. (2019). 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nature Biotechnology, 37*(2), 179-185. doi:10.1038/s41587-018-0008-8

Appendices

List of appendices:

Appendix A

BLISS gut microbiota questionnaire

The BLISS gut microbiota questionnaire was developed and collected at 7 and 12 months by the BLISS study team. The Candidate was responsible for entering all the data from the questionnaires into EXCEL.

Office use only

Date:

Code:

7 month Questionnaire - Microbiota

Thank you for your time.

1 Was your baby born by Caesarian?

- ○ No
- ○ Yes – Elective Caesarian (you knew the date and time beforehand)
- ○ Yes – Emergency Caesarian

2 Has your baby had any antibiotics in the past month?

- ○ No
- ○ Yes – and still having them
- ○ Yes – but has now stopped. Please state how long ago they stopped ……

3 Does your baby eat yoghurt?

- ○ No
- ○ Yes

If yes, how often do they eat yoghurt?

- ○ Less than once a month
- ○ Every month, but not every week
- ○ Every week, but not every day
- ○ Once a day
- ○ More than once a day

4 Do you have a freezer?

- ○ No
- ○ Yes – an icebox (a small box *inside* the fridge)
- ○ Yes – a fridge-freezer
- ○ Yes – a chest freezer
- ○ Yes – another type of freezer. Please describe ……

Thank you for completing this questionnaire

Office use only

Date:

Code:

12 month Questionnaire - Microbiota

Thank you for your time.

1 Has your baby had any antibiotics in the past month?

- ○ No
- ○ Yes – and still having them
- ○ Yes – but has now stopped. Please state how long ago they stopped ……

2 Does your baby eat yoghurt?

- ○ No
- ○ Yes

If yes, how often do they eat yoghurt?

- ○ Less than once a month
- ○ Every month, but not every week
- ○ Every week, but not every day
- ○ Once a day
- ○ More than once a day

3 Do you have a freezer?

- ○ No
- ○ Yes – an icebox (a small box *inside* the fridge)
- ○ Yes – a fridge-freezer
- ○ Yes – a chest freezer
- ○ Yes – another type of freezer. Please describe ……

Thank you for completing this questionnaire

Appendix B

BLISS 3-day weighed diet record

The BLISS 3-day weighed diet record was developed and collected at 7 and 12 months by the BLISS study team. The Candidate was responsible for checking and regrouping individual food items to the food groups used in this thesis.

BLISS Food Diary







Please read through the instruction pages before starting your food diary

Things to record each day:

	What	When
1	Food Diary	During preparation, dishing and clean-up of the food or drink
2	Description of Recipes Used	As you are cooking the recipe
3	End of Day Questionnaire	At the end of each day
4	Supplement Use	At the end of each day

On these days:

1. .....
2. .....
3. .....

Please try not to change what you give your child just because you are keeping a diary!

Thank you very much for your help

bliss

Day 1

Step 1	Step 2			Step 3		Step 4				Step 5		Step 6	
Time of day	Name of food or drink	Brand of food or drink	Cooking method	Weight of plate/ mug	Weight of food/drink + plate/ mug	Consistency of food/drink item				Food/drink was put in child's mouth by:		Weight of leftover + plate or mug	Estimation of what is left on plate
						Puréed	Mashed	Diced	Whole	Parent	Child		

If your child has finished eating for the day, please remember to fill out the end of day questionnaire and supplement use on pages 11 and 12.

B) How to describe recipes

Example:

Step 1	Step 2	Step 3	Step 4	Step 5
Name of recipe	Amount of each ingredient including any water added (eg: 3 medium carrots, 500g lean beef mince, 1 onion, 60g water etc)	Cooking method	Proportion of recipe served to your child e.g., one tenth or 1/10 or 10%	Time of day served

End of Day Questionnaire - Day 1

Table 1

Step 1	Step 2	Step 3		Step 4				
Date	Day of week	Is this a typical eating day for your child?		Is your child unwell (for any reason)?		If unwell, did this influence your child's appetite?		
		Yes	No	Yes	No	No	Yes - Decreased appetite	Yes - Increased appetite

Table 2

How did your child's meals compare to the family meals today?

	Part 1			Part 2										
				Step 1	Step 2		Step 3				Step 4			
	Did not have this meal today	Was with another adult at this meal time	Breast milk or formula **only**	Breast milk/ formula and food	Child ate meal with at least one other adult: (Both were eating but food may be different.) Yes	No	Meal Ingredients Exactly the same	Almost the same	Similar	Mostly different	Meal Preparation (Eg: texture of meal, length and method of cooking, size of food pieces) Exactly the same	Almost the same	Similar	Mostly different
Breakfast														
Lunch														
Evening meal														

Supplement Use – Day 1

(a) Did your child take any supplements today? Include anything you consider to be a supplement to your child's diet (e.g., multi-vitamin, etc.).

No ☐ (please go to page 13)

Yes ☐

(b) **If yes,** please record the following:

Type of supplement (e.g., cod liver oil): _____

Brand name (e.g., Smith's): _____

Amount (number of mls, drops, tablets, capsules, etc.) taken (e.g., 5mls): _____

(c) **If yes,** does the supplement contain iron or zinc? (check the label)

No ☐

Yes ☐

If yes, please record the type of iron (e.g., ferrous fumerate, ferrous sulphate and anything else with the words "iron", "ferric" or "ferrous" or "zinc" (e.g., zinc sulfate) and the amount of iron or zinc per tablet (e.g., 10mg, etc.):

Type of iron (e.g., ferrous sulphate): _____ Amount per dose (e.g., 7mg in 5ml): _____

Type of zinc (e.g., zinc sulfate): _____ Amount per dose (e.g., 7mg in 5ml): _____

THE INTERVIEWER WILL HELP YOU FILL IN THIS PAGE IF YOU ARE NOT SURE - please keep the bottle or packet

Appendix C

Energy contribution of individual food items to food groups (an example)

The Candidate was responsible for checking and regrouping individual food items to the food groups that were used in this thesis. This was done by grouping the 1682 individual food items (rows) into 18 food groups (columns). The snapshot of the Excel spreadsheet (Microsoft Excel, 2010) shows 11 out of 1682 individual food items (rows) with data that were added to indicate contribution of the food to energy intake from each of the 18 food groups (columns). For example: “Biscuit,semi-sweet” was categorised under the “Sweet food” food group so contributes 1894 kJ per 100 g to the “Sweet food” food group, but no energy to the “Breast milk” or other food groups.

descriptor	foodid	Breast_g	Breast_kj	Inftfor_g	Inftfor_kj	Cows_g	Cows_kj	Bread_g	Bread_kj	Inftcer_g	Inftcer_kj	Fruit_g	Fruit_kj	Vege_g	Vege_kj	Redme_g	Redme_kj	Fishpo_g	Fishpo_kj
Biscuit,chocolate coated	A1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Biscuit,semi-sweet	A10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bread,white,sliced,prepacked	A1007	0	0	0	0	0	0	100	961	0	0	0	0	0	0	0	0	0	0
Bread,wheatmeal,sliced,prepacked	A1008	0	0	0	0	0	0	100	894	0	0	0	0	0	0	0	0	0	0
Bread,multi-grain,light,sliced,prepacked	A1009	0	0	0	0	0	0	100	949	0	0	0	0	0	0	0	0	0	0
Croissant	A101	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bread,multi-grain,heavy,prepacked	A1010	0	0	0	0	0	0	100	939	0	0	0	0	0	0	0	0	0	0
Bread,ancient grains,vogels	A1024	0	0	0	0	0	0	100	860	0	0	0	0	0	0	0	0	0	0
Crispbread,corn,original,arnotts	A1029	0	0	0	0	0	0	100	1404	0	0	0	0	0	0	0	0	0	0
Cruskits,lite,arnotts	A1031	0	0	0	0	0	0	100	1390	0	0	0	0	0	0	0	0	0	0
Crackers,cream,reduced fat,huntley & palmers	A1033	0	0	0	0	0	0	100	1454	0	0	0	0	0	0	0	0	0	0

descriptor	foodid	Otherme_g	Otherme_kj	Legum_g	Legum_kj	Eggs_g	Eggs_kj	Nutse_g	Nutse_kj	Dairy_g	Dairy_kj	Beve_g	Beve_kj	Misc_g	Misc_kj	Swe_g	Swe_kj	Sav_g	Sav_kj
Biscuit,chocolate coated	A1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	2172	0	0
Biscuit,semi-sweet	A10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	1894	0	0
Bread,white,sliced,prepacked	A1007	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bread,wheatmeal,sliced,prepacked	A1008	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bread,multi-grain,light,sliced,prepacked	A1009	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Croissant	A101	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	1650
Bread,multi-grain,heavy,prepacked	A1010	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bread,ancient grains,vogels	A1024	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Crispbread,corn,original,arnotts	A1029	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cruskits,lite,arnotts	A1031	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Crackers,cream,reduced fat,huntley & palmers	A1033	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix D

Faecal sample collection procedures

The BLISS gut microbiota faecal sample procedures were developed by the BLISS study team.

If your child is using a nappy or training pants

1. After s/he has pooped in the nappy or training pants, collect the sample directly. Put on your gloves. Open the specimen jar. Using the scoop in the lid of the specimen jar, collect a small amount of poo (about the size of a 20 cent coin). Please try to avoid "wee" (urine) coming into contact with the poo.

2. Place the scooped poo sample into the specimen jar and firmly screw on the lid taking care that it is flat and shuts tight (is not cross-threaded). If necessary, please wipe the outside of the specimen jar so that it is clean.

3. Dispose of any remaining poo as you would normally. Put your gloves in the rubbish. Wash your hands well.

4. Please write the date the sample was collected on the specimen jar.

5. Put the sealed specimen jar in the small plastic bag provided and put this inside the blue and white flask.

6. Put the blue and white flask into your freezer immediately.

Pick up of Poo Sample

Once you have collected the poo sample and put it in the freezer, please let us know using any of the following methods:

Text your BLISS number and 'collect sample' to 0221 927 421
Email bliss@otago.ac.nz

A BLISS study researcher will contact you to arrange the collection.
Once you have put the sample in the freezer it can remain there for up to one week. We plan to drive around and collect the poo samples twice a week during the early evening hours. If this does not suit you we can arrange to collect it at another time. If you do not want to store the sample in your freezer, please let us know and we can arrange a more direct collection.
If you have any questions or would like to talk to someone about the collection please do not hesitate to contact us on 471 6063.

Instructions

We would like you to collect a small amount of your child's "poo" to look at how different babies' guts (large intestines) grow different amounts of good "bugs". We only need a very small amount – the size of a 20 cent coin is enough. There are many ways of collecting a poo sample but here are some tips to help you out. Remember not to make a big deal of it with your child.

- We all have bugs on our hands, but we need to make sure that they don't get into the sample. Therefore, we have provided a pair of disposable gloves for you to wear while collecting the poo. If you would rather just wash your hands with soap and warm water before collecting it, that's fine.

You will have been given:

- A brown specimen jar with a scoop attached under the lid
- A small plastic bag
- One pair of disposable gloves (and a spare pair just in case!)
- A blue and white insulated flask

If your child uses the toilet (rather than a potty) you will also have been given:

- A container to 'catch' the sample
- A large bag for disposal of the container

Collecting the Specimen

It would be a good idea to keep the BLISS poo collection pack in the room where your child usually does their poos, out of the reach of children, so that it is ready for you to use as soon as there is an opportunity.

bliss

Thank you very much. We know that a poo sample is an unusual request!
A poo sample is the best way to find out what bugs are naturally in babies' guts – and how this is related to what they eat.

If your child usually uses the toilet

1. When s/he is ready to do poos, put on your gloves, place the BLISS-poo container flat in the toilet bowl to collect the poo – the container should sit above the water line. Do not collect the poo from the toilet bowl.

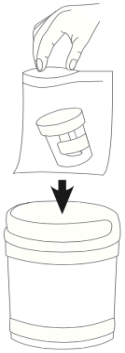
2. After s/he has done poos inside the BLISS-poo container, remove the container from the toilet and open the specimen jar. Using the scoop in the lid of the specimen jar, collect a small amount of poo (about the size of a 20 cent coin). Please try to avoid "wee" (urine) coming into contact with the poo (we know this can be very difficult to do but would really appreciate it if you could try).

3. Place the scooped poo sample in the specimen jar and firmly screw on the lid taking care that it is flat and shuts tight (is not cross-threaded). If necessary, please wipe the outside of the specimen jar so that it is clean.

4. Flush any remaining poo and wee down the toilet and place the BLISS-poo container you used to 'catch' the poo, and your gloves, into the large plastic bag provided and dispose of it in the rubbish. Wash your hands well.

5. Please write the date the sample was collected on the specimen jar.

6. Put the sealed specimen jar in the small plastic bag provided and put this inside the blue and white flask.



7. Put the blue and white flask into your freezer immediately.

If your child usually uses a potty

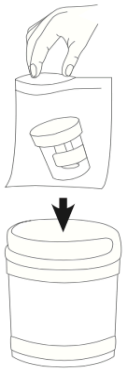
1. After s/he has pooped in the potty, collect the sample directly from the potty. Put on your gloves. Open the specimen jar. Using the scoop in the lid of the specimen jar, collect a small amount of poo (about the size of a 20 cent coin). Please try to avoid "wee" (urine) coming into contact with the poo (we know this can be very difficult to do but would really appreciate it if you could try).

2. Place the scooped poo sample into the specimen jar and firmly screw on the lid taking care that it is flat and shuts tight (is not cross-threaded). If necessary, please wipe the outside of the specimen jar so that it is clean.

3. Dispose of any remaining poo and urine as you would normally. Put your gloves in the rubbish. Wash your hands well.

4. Please write the date the sample was collected on the specimen jar.

5. Put the sealed specimen jar in the small plastic bag provided and put this inside the blue and white flask.



6. Put the blue and white flask into your freezer immediately.

Appendix E

Statistical analysis plan & lessons learnt for study 1

The Candidate developed the statistical analysis plan with advice from supervisors JH, ALH and RT. References are not included as they can be found in the main body of the thesis.

Statistical analysis plan for BLISS microbiota paper

Version: 8 Feb 2017

Title: Does a baby-led approach to feeding solids affect the infant gut microbiota?

What we know:

- Rapid growth of microbiota colonies in the first few years of life.
- The introduction of complementary foods changes the gut microbiota drastically.
- The microbiota developed in early life is very important for immune function, gut-brain axis etc.
- Most microbiota changes occur in infant life with less microbiota changes after 3 years of life.
- Mode of delivery, use of antibiotics, breastfeeding and formula feeding affects the microbiota.
- The diet affects the microbiota.

BLISS	Control (current MoH complementary feeding guidelines)
Differences	
Self-feed	Care givers feed
Uses whole foods	Start with pureed food, then progress to mashed -> chopped -> whole
Share family food from start	Eat family food at a later age
Proposed benefits	
More varied food intake	Less risk of iron deficiency
Improved appetite control	Less risk of growth faltering
Mess (less hygiene)	Care giver can choose foods and amount for infants
Start complementary food later (44% ate any ready to eat commercial food)	
Have better fine motor skills	

What we want to know (Research questions):

1. Does a baby-led approach to complementary feeding (BLISS-approach) impact on the microbiota at 7 months of age, or 12 months of age?
2. Does a baby-led approach to complementary feeding (BLISS-approach) impact on the change in microbiota between 7 and 12 months?
3. Does a complementary diet with more high-fibre foods, or a greater variety of fibre sources, predict microbiota at 7 and 12 months?

In other words:

- Will the BLISS approach have a microbiota more towards an adult microbiota at an earlier age as they share family food from the start?

How to find out the answers:

Study design (Figure 1)
-Methods

Table 1: Study demographics of BLISS vs Control (Descriptive data of the study groups)

- gestational age
- delivery mode
- infant sex
- infant birth weight
- feeding at 7 months (breast milk as only infant milk, combination feeding, infant formula as only infant milk)*
- age of formula introduction of combination fed infants
- age of introduction of solids
- recent antibiotic use
- parity

Figure 2: Alpha diversity plots [observed OTUs against no. of sequences] or Shannon index. Comparisons between BLISS and Control at both 7 and 12 months and also between 7 and 12 months (all four trajectories/box plots on one plot, with 95% CI at the final (30000?) number of sequences).

Table 2: Relative abundance of the 5-10 most abundant bacterial family. Difference between BLISS and control at both 7 months and 12 months and then at the change between 7 and 12 months.

➔ Adjust for confounders (Unadjusted model and adjusted model)

1. Current breastfeeding status (breast milk as only infant milk, combination feeding, infant formula as only infant milk)*
2. Mode of delivery [if literature suggests is still relevant at 7mos]
3. Parity
4. Infant antibiotics used in the past month [if sufficient cases]

Table 3: Predictors table (adjusted and unadjusted)

➔ If no differences between groups can combine the groups
➔ Predictors to look at

1. Current breastfeeding status (breast milk as only infant milk, combination feeding, infant formula as only infant milk)*
2. Age of introduction of complementary foods
3. Dietary fibre intake (g)
4. Fibre variety score (sources of grains/fruits/vegetables)
5. Fruit/ vegetable intake (not incl fruit juice) (g)
6. Ever eaten legumes
7. Ever had infant formula
8. Age when infant cereal (fortified with iron) was first eaten

➔ Confounders to put into model
1. Group (BLISS vs Control)
2. Mode of delivery [if literature suggests is still relevant at 7mos]
3. Parity
4. Infant antibiotics used in the past month [if sufficient cases]

➔ Interesting to consider:
1. Infant fine motor skills
2. Whole foods (BLISS) vs pureed foods (control)
3. Total/ other food variety scores

Figure 3: Figure looking at changes of proportion of fruit/vegetable/breast milk/formula milk intake together with changes in microbiota (relative abundance or alpha diversity depending on what comes up in the earlier analyses)

NOTE:
*Get this from our weeks of breastfeeding and infant formula feeding data from the questionnaires, and check this against the WDR to confirm behaviour.

Why are these predictors & confounders chosen?

No.	Factor	Relation to microbiota	Age group	Reference
1	Current breastfeeding status	Lower bacterial richness and *alpha* diversity in infants who were breastfed.	4 mth	(Azad et al., 2013)
		Microbiota of the breast-fed infant appears to develop more slowly than that of the formula-fed infant for the fermentation of complex carbohydrates.	[Review]	(Edwards & Parrett, 2002)
		The pre-weaning feeding method (breast-fed, formula-fed or mixed-fed) influenced the relative proportions of Bacteroides, members of the genus Bifidobacterium.	4 wks after weaning	(Fallani et al., 2011)
2	Mode of delivery	Caesarean was associated with a lower *alpha* microbial diversity.	24 mth	(Jakobsson et al., 2014)
		Differences in species diversity between delivery modes are decreased after 4 months, and almost disappear by 12 months of age.	4 & 12 mth	(Bäckhed et al., 2015)
3	Parity	Parity was not associated with differences in the breast milk microbiota or with changes in the infant faecal microbiota.	< 12 mth	(González et al., 2013)

		Parity <-> Breast milk <-> IgA <-> Microbiota	4 mth	(Bridgman et al., 2016)
4	Infant antibiotic use	Antibiotic treatment changes the microbiota.	[Review]	(Willing, Russell, & Finlay, 2011)
		Antibiotic use may cause a shift in microbiota.	Adults	(Dethlefsen & Relman, 2011)
5	Age of introduction of complementary foods	Age of introduction of complementary foods did not have any impact on the microbiota differences after correcting for pre-weaning feeding method.	4 wks after weaning	(Fallani et al., 2011)
		Age of introduction to complementary feeding was generally not correlated with alpha diversity measures.	9 mth	(Laursen et al., 2016)
6	Dietary fibre intake	There are differences in microbiota associated with long-term ingestion of a diet rich in fruit and vegetable fibre.	[Review]	(Simpson & Campbell, 2015)
7	Fibre variety score	An association between microbiota richness and diversity of vegetables the participants were eating before the trial was observed.	Adults	(Tap et al., 2015)
8	Fruit/ vegetable intake (not incl fruit juice) (g)	Food components, which are indigestible for human enzymes (e.g. fibre), provide substrates for the microbiota metabolism.	[Review]	(Graf et al., 2015)
9	Ever eaten legumes	Long-term intake of a diet that is high in fruit and legume fibre (agrarian diet) is associated with greater microbiota diversity.	1-6 yr	(De Filippo et al., 2010)
10	Ever had infant formula	Introduction of formula milk to breastfed infants led to a microbiota that is more similar to those that were fully formula fed.	6 wks	(Madan et al., 2016)
11	Whole foods (BLISS) vs pureed foods (control)	Lack of chewing ability and immature pancreatic exocrine function in infants may lead to more starch escaping digestion.	[Review]	(Edwards & Parrett, 2002)

Lessons learnt for the Candidate:

- As the first statistical analysis plan developed, this plan helped the Candidate realise the importance of coming up with an analysis plan, especially in studies that are secondary analyses. It helps to gather the knowledge in the literature, the research gaps, what data are available and plan ways to analyse the available data in a systematic manner to answer the research questions raised.
- The study design is important. As the study was a randomised controlled trial (RCT), and had longitudinal data, statistical analysis such as mediation analyses could be used.
- The sample size is important. The sample size of 74 is larger than many other gut microbiota studies. However, it is still not large enough to be suitable for statistical analyses such principal component analysis (PCA). In addition, the food groups developed had to be limited to 9 food groups instead of the 18 initially developed due to having sufficient number of participants (i.e. at least 8 participants) for each food group tested.
- It is important to come up with initial research questions. As the gut microbiota dataset is huge, there can be many different indices of the gut microbiota to analyse (e.g. different alpha diversity measures, hundreds of family-level bacterial taxa). It is important to look at the effect sizes and not base conclusions on p-values alone.
- It is also important to consider confounders. In human studies, there are many confounders that may affect the analyses and need to be considered. Hence, a good literature review of the confounders needs to be developed and used in the analyses where appropriate.
- Choosing the food groups to be developed. As there is a limit to the number of food groups to be developed, it is important to include rationale on how and why the individual food items from the weighed diet records are grouped together.
- Mediation models are important in the gut microbiota area as they are able to assess causal hypotheses. However, when using the mediation models, it is important to ensure that the data meet the criteria for mediation.

Appendix F

Statistical analysis plan & lessons learnt for study 2

The Candidate developed the statistical analysis plan with advice from supervisors JH, ALH and RT. References are not included as they can be found in the main body of the thesis.

Statistical analysis plan for EAT5 validation paper

Version: 8 May 2018

Title: Relative validity and reproducibility in pre-schoolers of an FFQ for determining intake of nutrients and food groups of relevance to the gut microbiota

Aim: To determine the ability of a 123-item food frequency questionnaire to assess the usual dietary intake over the past month of New Zealand pre-schoolers aged 5 years old.

Objectives:
To determine the relative validity and reproducibility of the EAT5 FFQ for assessing:
Primary:

- amount and ranking of dietary fibre and fibre fraction intake.

Secondary:

- amount and ranking of food group intake.
- amount and ranking of energy, protein, carbohydrate, fat, saturated fat, total sugars, iron, vitamin C and calcium intake.

What we know:

- Nutrient, food groups of relevance to the gut microbiota:
    - Whole grains (Cooper et al., 2017)
    - Nuts (Lamuel-Raventos & Onge, 2017)
    - Legumes (Fernando et al., 2010)
    - Resistant starch (Mitsou et al., 2011)
    - Pectin (Koutsos, Tuohy, & Lovegrove, 2015)
    - Polyphenols (Duenas et al., 2015)-review
    - Soymilk (Fujisawa, Ohashi, Shin, Narai-Kanayama, & Nakagaki, 2017)
    - Prebiotics (Holscher, 2017)-review, (Louis, Flint, & Michel, 2016)-review
    - Oats (beta-glucan) (Rose, 2014)-review
    - Probiotics (Singh et al., 2017)-review
- Validation and reproducibility of a number of FFQs available in young children (with dietary fibre included in validation): (Moghames et al., 2016)-FFQ compared against four 24hour recalls. (Vioque et al., 2016)-FFQ compared against three 24hour recalls. (Watson et al., 2015)-FFQ compared to 5day WDR.
- Validation and reproducibility of a number of FFQs available in young children (with fruit and vegetables included in validation): (Flood et al., 2013; Huybrechts, De Backer, De Bacquer, Maes, & De Henauw, 2009; Mills et al., 2015)
- Validation of FFQ specially looking at dietary fibre (Healey et al., 2016; Reeves, Winkler, & Eakin, 2015).
- Validation of FFQ specially looking at inulin and oligosaccharides in adults (Dunn et al., 2011).
- Importance of the gut microbiota in young children: There is a critical window in early life for influencing health as the gut microbiota is still establishing till about 3 to 5 years of age (Cheng et al., 2015). Moreover, childhood obesity has been associated with increased Firmicutes-to-Bacteroidetes ratio (Bervoets et

al., 2013) and higher concentration of *Enterobacteriaceae* (Karlsson et al., 2012).

- Definitions:
    - Whole grains: "Wholegrain food is any food which uses every part of the grain including the outer layers, bran and germ. This definition applies even if these parts are separated during processing and regardless of whether the grain is in one piece or milled into smaller pieces. Under the Food Standards Code Standard 2.1.1 the term wholegrain refers to: the term wholegrain refers to: -Whole and intact grains as found in some bread and crisp breads. -Puffed or flaked grains in some breakfast cereals. Coarsely milled or kibbled wheat found in breads such as pumpernickel. -Ground grains such whole wheat flour used to make wholemeal bread. The term wholemeal applies to foods in which the whole grains have been refined into finer particles. This gives manufacturers the option of describing their foods as either wholegrain or wholemeal to avoid misleading the customer." (Food Standards Australia New Zealand, 2018)
    - Prebiotics: "Selectively fermented ingredients that allow specific changes, both in the composition and/or activity in the GI microflora that confer benefits upon host wellbeing and health." The three criteria required for a prebiotic effect are as follows: 1. Resistant to gastric acidity and hydrolysis by mammalian enzymes and GI absorption. 2. Can be fermented by intestinal microflora. 3. Selectively stimulates the growth and/or activity of intestinal bacteria associated with health and wellbeing. (Gibson, Probert, Loo, Rastall, & Roberfroid, 2004)
    - Probiotic: "Live microorganisms that, when administered in adequate amounts, confer a health benefit on the host."
    - Dietary fibre: "Dietary fibre means carbohydrate polymers with ten or more monomeric units, which are not hydrolysed by the endogenous enzymes in the small intestine of humans and belong to the following categories: 1. Edible carbohydrate polymers naturally occurring in the food as consumed. 2. Carbohydrate polymers, which have been obtained from food raw material by physical, enzymatic or chemical means and which have been shown to have a physiological effect of benefit to health as demonstrated by generally accepted scientific evidence to competent authorities. 3. Synthetic carbohydrate polymers, which have been shown to have a physiological effect of benefit to health as demonstrated by generally, accepted scientific evidence to competent authorities. In the footnote: The decision on whether to include carbohydrates from 3 to 9 monomeric units should be left to national authorities." (CODEX, 2009)

Steps taken to get data for analysis:

- Data entry of FFQs n=25*2
- Matched individual foods (n=1010) from the diet records to fibre and fibre fractions in Kaiculator.
- Matched individual foods (n=1010) from the diet records to food groups (n=34) in Kaiculator. Regrouped 34 food groups to 12 food groups.

How to find out the answers:

Table 1: Demographics.

Table 2: Food groupings used for the EAT5 FFQ (n=12)

Table 3: Average daily intake of selected nutrients FFQ1 and diet record, both crude and FV (Fruit and vegetable) adjusted values among children aged 5 years (n=100).

Table 4: Relative validity of FFQ1 compared with the diet record.
(First look at normality by visual checks of histograms. Then, compare FFQ1 and diet record selected nutrients and food groups using paired t-test, cross-classification (correct + gross misclassification), and Spearman's correlation coefficients and ICC.)

Table 5: Reproducibility of the FFQ in terms of selected nutrients, food components and food groups (n=99).
(Using paired t-test and intra-class correlation coefficients)

Figure 1 (Or maybe supplementary or text or thesis): Bland-Altman plot for dietary fibre, pectin, resistant starch etc.

Table 6 (Or maybe supplementary or text or thesis): Bland-Altman statistics comparing nutrient intakes from FFQ1 and diet record for both crude and FV-adjusted values.
(Mean % agreement, 95%CI, Limits of agreement).

NB: Remember to consider de-attenuated values for validity correlations.

To note about missing amount data from FFQs:

For relative validity: Use imputed values.
Conduct sensitivity analyses using completed data (<5% missing of 123 food items) – Come up with ID of complete vs incomplete data.
For reproducibility: Use only the completed data (<5% missing of 123 food items).

N.B. It is ok to impute in real world as FFQ should not be used for individuals.
This is a conservative solution as we are using group value rather than actual food diary amount.

Lessons learnt for the Candidate:

- Compared to study 1 and study 3, statistical analyses for study 2 were in a sense, more straightforward as the statistical techniques for validation studies are well established.
- It is important to have a codebook for data entry. There were food items in the FFQs where participants had indicated serving sizes that were different to the units of measurement provided. The way to address this problem was to develop a codebook for data entry and reuse the same units or volumes for the problematic food items across all the FFQs.
- Computation of missing data. There were a few food items in the FFQs that had frequencies but had missing amounts. As the data were collected before the Candidate started her studies, the reasons for the missing data could not be investigated and imputation of missing values had to be undertaken.
- Similar to study 1, there was the challenge of choosing the food groups to be used. As there is a limit to the number of food groups that can be used, it is important to have a rationale for how and why the individual food items from the weighed diet records are grouped together.
- Choosing the statistical tests to undertake. There are many different types of statistics that can be used to determine the validity and reproducibility of an FFQ. It is important to consider the nature of the data and use a number of statistical methods to get an overview of how well the FFQ performs.
- Solving technological problems. Technology is great and can allow us to get the results efficiently. However, technology does come with its limitations. For example, it is important to check the script used in the coding for getting the nutrient values from the online FFQ data entry. This is because if the researcher is unable to get the script and just the outputs of the nutrient analyses, it is just like a 'black box' where you have to trust that the coding in the 'black box' is accurate.

Appendix G

EAT5 food frequency questionnaire

The EAT5 FFQ was developed by supervisors ALH, RT and Dr. Sonya Cameron. 99 of the EAT5 FFQs and 3-day WDRs were collected by 5 Masters of Dietetics students from the department of Human Nutrition. The Candidate collected 1 set of EAT5 FFQs and 3-day WDR; keyed in all 199 EAT5 FFQs and checked all 100 WDRs. The FFQ collected at 5 years for the POI study was the same FFQ as the EAT5 FFQ, with a different study name and contact details. The POI5 FFQs were collected by the POI study team and the Candidate was responsible for entering data from all 546 POI5 FFQs into the online FFQ database that was specifically developed for the EAT5 and POI projects.

5 year Food Frequency Questionnaire

To be administered to the child's main carer

EAT5 study number

Interview date DD MM YY

Interviewer name _____

You will need:

- This questionnaire
- Ready reckoners 1 and 2
- Plate
- Bowl
- Mug
- Glass
- Measuring cup
- Measuring spoons
- Funnel
- Beans
- Rice

Instructions for the interviewer:

- Please check that the POI study number, the date, and your name are **recorded on the front page** before starting the questionnaire
- Ask each question in turn **as it is worded** in the questionnaire.
- Each food should be **entered only once** and **as separate items** where possible. For example, if they had custard and peaches for pudding, this should be entered as i) custard and ii) peaches – not as "puddings not yet described".
- **Mixed dishes should be entered separately** – so lasagne should be entered as appropriate amounts of "Other pasta (pasta only - not including sauce)", "Mince & patties (from beef or lamb)", and "White or cheese sauce", and "Cooked tomato (pasta sauce, canned tomatoes etc)".
- Room has been left at the bottom of each page to **make comments** that will help with interpreting the data – we would like you to put as much information as possible. For instance a child might have a slice of bread that is from a breadmaker which is much thicker than the standard toaster slice.

Choosing a frequency

- Enter **only one value for frequency** (that is, how often a week OR day, not both)
- Make sure there is a frequency for **every food** (i.e., use the "not eaten this month" column as appropriate)
- Use "Ready Reckoner 1" if parent says that the child has something **regularly but only for part of the week,** for example, "Beth has a banana twice a day during the week but not at all in the weekend". This will give you the single frequency option to use.
- Use "Ready Reckoner 2" for those **questions that include more than one food,** for example "white buns (not iced), crumpets". This will give you the single frequency option to use.

Estimating the amount eaten

- **For those foods that have a "standard" unit provided,** the parent needs to estimate how many of these units (or fractions of units such as ¼ or ½) the child would usually eat (1st example on the next page).
- **For those foods with units of ml,** you need to use the beans (larger foods) or rice (smaller foods) to estimate the volume they consume. Ask the parent to pour the beans or rice into the plate or cup (whichever is appropriate) to the level that their child would usually eat. You then pour this amount into the measuring cup and record the volume (2nd example on the next page).

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
Example	**Food**	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
1	Instant noodles			✓								1 packet	¾
2	Cornflakes or rice bubbles							✓				140 ml	

Instructions for the participant:

- These questions ask how often and how much *child's name* has eaten certain foods or beverages **over the past month.**
- *Child's name* may sometimes be fed by a relative, friend, or someone else. If you know the type of food and approximate amount s/he has eaten at these times please include them.
- Please tell us what they **actually** ate and drank; we are interested in what children actually eat, not the perfect diet.
- All information is stored by study number and not by your name.

A. Bread, crackers and breakfast cereals

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
1	White bread											1 toaster slice	
2	White buns (not iced), crumpets											1 small bun or 1 crumpet	
3	Wholemeal bread or bun											1 toaster slice or 1 small bun	
4	Wholegrain bread or bun											1 toaster slice or 1 small bun	
5	Crackers (wheat, rice or corn-based)											1 cracker	
6	Rice cakes or rice wheels*											1 rice cake or 17 rice wheels	
7	Cruskits or crispbreads											1 cruskit	

Comments?

* Rice wheel = tiny kids' size rice cake (~ 50 cent coin in diameter)

* Rice cake = puffed rice crackers the size of a large cookie

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
8	Weet-bix											1 weetbix	
9	Fruity-bix or similar											_____ ml	
10	Porridge											_____ ml	
11	Cornflakes or rice bubbles											_____ ml	
12	Cocopops, honey puffs or puffed wheat cereal											_____ ml	
13	Nutrigrain, Milo cereal or similar											_____ ml	
14	Muesli and "light" muesli (e.g. "Light and Tasty, Light and Right)"											_____ ml	
15	Other breakfast cereal (record name of main one): _____											_____ ml	

Comments?

B. Rice and pasta

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
16	White rice											_____ ml	
17	Brown rice											_____ ml	
18	Instant noodles											1 packet	
19	Canned spaghetti											1 medium can (420g)	
20	Other pasta (pasta only - not including sauce)											_____ ml	
21	Pizza (not takeaway) – base only											1 slice	

Comments?

C. Fruit

22 How often has your child had fruit in the past month?

_____ less than once a week OR _____ times a week OR _____ times a day

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
23	Banana											1 banana*	
24	Apples (fresh or canned)											1 apple	
25	Apricots, plums or peaches (fresh or canned)											1 apricot or 1 plum, or ½ peach	
26	Pears (fresh or canned)											1 pear	
27	Raisins or sultanas											1 small box (12g)	
28	Dried apricots or prunes											1 prune	
29	Oranges, mandarins											1 orange or 2 mandarins	

Comments?

*2 bobby bananas = 1 medium banana

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
30	Kiwifruit											1 kiwifruit	
31	Green grapes											1 grape	
32	Black or red grapes											1 grape	
33	Berries or cherries (fresh or frozen)											_____ ml	
34	Avocado											1 avocado	
35	Rhubarb											_____ ml	
36	Other fruit											_____ ml	

Comments?

D. Vegetables

37 How often has your child had vegetables in the past month?

_____ less than once a week OR _____ times a week OR _____ times a day

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
38	Potato salad or other potato eaten cold											_____ ml	
39	Potato or kumara (boiled, baked, microwaved, mashed) eaten warm or hot											_____ ml	
40	Hot chips, potato shapes, roast potato or kumara cooked at home eaten warm or hot											_____ ml	
41	Yams											_____ ml	
42	Carrot											_____ ml	
43	Pumpkin											_____ ml	
44	Green peas											_____ ml	
45	Green beans											_____ ml	

Comments?

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
46	Sweet corn											_____ ml	
47	Broccoli											_____ ml	
48	Cauliflower											_____ ml	
49	Capsicum (peppers)											_____ ml	
50	Red cabbage											_____ ml	
51	Green cabbage											_____ ml	
52	Spinach or silverbeet											_____ ml	
53	Lettuce or salad leaves											_____ ml	
54	Cucumber											1 slice	
55	Raw tomato											1 tomato	
56	Cooked tomato (pasta sauce, canned tomatoes, tomato sauce)											_____ ml	
57	Leeks											_____ ml	
58	Other vegetables: _____											_____ ml	

Comments?

E. Meat, chicken, fish, eggs, beans

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
59	Chicken nuggets or shapes											1 nugget	
60	Fish fingers or shapes											1 fish finger	
61	Battered or crumbed fish											1 fillet	
62	Other chicken (eg. roast, stir-fry, BBQ)											1 drumstick	
63	Other fish (eg. canned, pan-fried)											_____ ml	
64	Sausages, saveloys (including vegetarian)											1 sausage	
65	Ham, bacon, luncheon, salami											1 slice	
66	Meat pies											1 pie	
67	Sausage rolls											1 roll	
68	Mince & patties (from beef or lamb)											_____ ml	

Comments?

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
69	Steak, chops or roast (beef or lamb)											_____ ml	
70	Pork and other meat											_____ ml	
71	Eggs											1 egg	
72	Hummus (chickpea dip)											_____ ml	
73	Baked beans											1 medium can (420g)	
74	Canned or home cooked beans, chickpeas or lentils											_____ ml	
75	Nuts (any sort but not peanut butter)											_____ ml	

Comments?

F. Spreads

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
76	Jam or honey											_____ ml	
77	Marmite or Vegemite											_____ ml	
78	Peanut butter											_____ ml	
79	Nutella											_____ ml	

Comments?

G. Cakes, biscuits, snacks

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
80	Biscuits – chocolate coated											1 biscuit	
81	Biscuits – other											1 biscuit	
82	Cakes or slices											1 large raisin box	
83	Muffins or scones											1 medium muffin	
84	Croissant, sweet buns, iced buns, pastries											1 item	
85	Fruit bread, currant buns											1 slice	
86	Chocolate											1 square	
87	Lollies											1 lolly	
88	Crisps, corn chips, corn snacks (e.g. Cheezels)											1 snack packet (18g)	
89	Muesli, nut, cereal or puffed rice bars											1 bar	
90	Fruit leather, fruit strings, fruit roll-ups											_____ ml	

Comments?

H. Milk and dairy products

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
91	Flavoured milk (including Milo, Quick, Drinking chocolate, Up-and-Go)											_____ ml	
92	Low-fat cows milk (green, lite blue, yellow-top) **as a drink**											_____ ml	
93	Low-fat cows milk **on cereal or other food** (not custard or sauces)											_____ ml	
94	Cows milk (blue, silver-top) **as a drink**											_____ ml	
95	Cows milk **on cereal or other food** (not custard or sauces)											_____ ml	
96	Soy milk **as a drink**											_____ ml	
97	Soy milk **on cereal or other food** (including custard or sauces)											_____ ml	
98	Other milk (goat, rice) **as a drink**											_____ ml	
99	Other milk (goat, rice) **on cereal or other food**											_____ ml	

Comments?

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
100	Cheese (including in recipes)											_____ ml	
101	Yoghurt or dairy food											_____ ml	
102	White sauce or cheese sauce											_____ ml	
103	Butter (not in baking)											_____ ml	
104	Margarine (not in baking)											_____ ml	
105	Cream or sour cream											_____ ml	

Comments?

I. Puddings

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
106	Ice cream											_____ ml	
107	Custard and other milk puddings											_____ ml	
108	Puddings not yet described											_____ ml	

Comments?

J. Drinks

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
109	100% orange juice (freshly squeezed or similar)											_____ ml	
110	Other fruit juice ("Fresh up", "Just Juice")											_____ ml	
111	Fruit drinks, Ribena, cordial, sachets											_____ ml	
112	Regular fizzy drinks (lemonade, coke)											_____ ml	
113	Diet fizzy drinks (lemonade, coke)											_____ ml	
114	Tea (not herbal)											_____ ml	
115	Coffee											_____ ml	

Comments?

K. Takeaways

		How often have they eaten this food this month? *Complete one frequency column only per food*										How much did they eat each time?	
	Food	Not eaten this month	Less than once a week	Once a week	2 times a week	3 times a week	4 times a week	5 times a week	6 times a week	Every day	If more than once a day – how many times a day	Units	How many [units] would they eat each time?
116	Chips from a takeaway shop or fast food restaurant (KFC, McDonalds, Burger King)											_____ ml	
117	Hotdog, fish, sausage from a takeaway shop											1 item	
118	Burgers from a takeaway shop or fast food restaurant (KFC, McDonalds, Burger King)											1 burger	
119	Other item from a takeaway shop or fast food restaurant (KFC, McDonalds, Burger King (please describe)											1 item	
120	Ready to eat pizza (takeaway shop or supermarket)											1 slice	
121	KFC or other fried chicken											1 piece	
122	Subway sandwich											1 6" roll	
123	Kebabs or wraps (bought)											1 wrap	
124	Sushi (bought)											1 piece	
125	Chinese, Thai, or Indian meal or similar (bought)											_____ ml	

Comments?

126 On average over the past 4 weeks, how many meals **per week** were given to your child by someone other than yourself? _____ meals

127 If more than 0 meals, how many of these meals have you been able to include in this questionnaire?

☐ None ☐ Some ☐ Most ☐ All

Thank you for completing this questionnaire

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Department of Human Nutrition and
Department of Medicine
University of Otago
P.O. Box 56, Dunedin 9054
Email:
Phone:

eat5

Appendix H

Protocol for checking WDR data

The Candidate was responsible for developing the protocol, and used the protocol to check all 100 WDRs.

Protocol for checking weighed diet records	
Study: EAT5 Study	*Version number:* Version 1
Prepared by: Claudia Leong	*Date prepared:* 3 Jan 18
Revisions: 8 Jan 18, 9 Mar 18	*Final:* 16 Mar 18

Equipment required

- Access to Kai-culator.
- Access to EAT5 'Protocols' folder.
    a. WDR checking folder
    b. WDR EAT5 codebook
    c. WDR checking template
- Use a red coloured pen for all checking.
- Hardcopy of weighed diet record.

Calculation checking

1. Check that all foods from each day have been included in each day's calculation sheet.
2. Check that the correct data from the weighed diet record have been entered.
3. Check each calculation for any inaccuracies.
4. Corrections should be made on the hardcopy in red pen.

Weighed diet record checking

1. Open Kai-culator and select the correct project.
2. Select the weighed diet record that you wish to check.
3. For each food in the hardcopy, check that it matches the data in the weighed diet record in Kai-culator.
    a. Check appropriateness of the food item
    b. Check completeness of the food item
    c. Check that there is no double entry
    d. Check that the recipes recorded within the diary are entered correctly
    e. Check that those generic food items are standardized by referring to the 'WDR EAT5 codebook'
4. Any revisions to the data should be made in red pen.
5. Before making any changes in Kai-culator, fill in the 'WDR checking template' found in the 'Protocols' folder on Dropbox.
6. Save any changes made.
7. In Kai-culator, double click on the weighed diet record if you wish to make changes.
8. To change the 'descriptor', click on the item in the 'diary item' column. Click SAVE.
9. To change the 'time', 'food item', or 'amount', click on the item in the relevant column. Click SAVE.

10. When all changes have been made, save and exit the weighed diet record.

Recipes

If the weighed diet record does not include all the recipes within the diary also check the Recipe database:

Before making any changes in Kai-culator, fill in the 'WDR checking template' found in the ‘Protocols’ folder on Dropbox.

1. Individually check each recipe entered in the Recipe Database.
2. Go to Recipes in Kai-culator, find the name of the recipe, open recipe (Click on ‘Edit recipe’ icon).
3. Check that each ingredient, food weight and cooking method have been correctly entered.
4. To change the ‘food item’ or ‘amount’, click on the item in the relevant column. Click SAVE.
5. When all changes have been made, save and exit the recipe.
6. Remember to click SAVE on the overall Recipe Database page to ensure all changes have been saved.

WDR checking template

Column headings:

	A	B	C	D	E	F
1	ID	Day	Changes required	Descriptor	Food item	Revised food

G	H	I
Amount	Revised amount	Date modified

Appendix I

Codebook for entering FFQ data

The Candidate was responsible for developing the codebook, and used the codebook when entering all 199 EAT5 FFQs and 546 POI5 FFQs.

POI 5 year FFQ codebook_6Oct2017		
Question number	**Item**	**Details**
33	Berries	1 Strawberry = 15 mL
34	Avocado	1 slice = 0.125
36	Other fruit	Feijoa = 30 mL
40	Hot chips	1 chip = 15 mL
42	Carrot	1 carrot = 250 mL; 1 baby = 30 mL
45	Green beans	1 bean = 5 mL
46	Sweet corn	1 sweet corn = 250 mL
47	Broccoli	1 floret = 15 mL
48	Cauliflower	1 floret = 15 mL
49	Capsicum	Capsicum = 220 mL; 1 slice = 30 mL
52	Lettuce	1 leaf = 15 mL
55	Raw tomato	1 whole = 127g ; 1 cherry/ slice = 20g
63	Other fish	1 fillet = 150 mL
64	Sausages	1 cheerio = 0.5 sausage
68	Mince	1 patty = 50 mL
69	Steak	1 chop = 150 mL; 1 palm = 75 mL
75	Nuts	1 handful = 30 mL
90	Fruit leather	1 fruit leather = 28 mL
100	Cheese	1 small rasin box = 12 mL; 1 slice = 15 mL
101	Yoghurt	1 pottle = 125 mL
106	Ice cream	1 scoop = 50 mL
108	Other pudding	1 standard ice block = 90 mL; mini ice block = 45 mL
116	Chips	1 small packet = 150 mL
119	Other fast food	To place nuggets in 121 (KFC chicken); ice cream sundaes and milkshakes in 106 (ice cream); pineapple ring in 36 (other fruit)
121	KFC chicken	1 piece fried chicken = 68g; nugget = 18g; chicken bite = 5g
122	Subway	1 6 inch roll = 220g; 1 kiddy sandwich = 130g
124	Sushi	1 sushi ball = 5 pieces
Other		
For other food items, Qn 15, 36, 58, 119, if what is written can be found in the list of items, to be placed back into the correct group.		
If have value (amount) but no frequency -> put frequency as 0.5		
If have frequency but no value -> put down frequency and leave value (amount) as BLANK		
If have 2 different frequencies, use the lower one or the middle whole value.		
1 cup = 250 mL		
Remember to calculate to the unit given: eg. Qn 6: 25 rice wheels = 25/17 = 1.5 unit		
If food item is given in grams and unit instead of mL: Look at Kaiculator, take average grams of the item and calculate mL from form density		
If have a range for the amount, use the middle value.		
Handful = 2		

Appendix J

Protocol for checking FFQ data

The Candidate was responsible for developing the protocol, and used the protocol to check all 199 EAT5 FFQs and 546 POI5 FFQs.

Protocol for checking and keying in Food Frequency Questionnaires

Study:	EAT5 Study	*Version number:* Version 2
Prepared by:	Claudia Leong	*Date prepared:* 6 Oct 17
		Date revised: 15 Jun 18

Equipment required

- Access to online FFQ database submission 'https://ffq.otago.ac.nz/ffq5yrs'.
- Access to POI5 'Protocols' folder.
    a. FFQ checking folder
    b. POI5 FFQ codebook_6Oct17
- Use a red coloured pen for all checking and data entry.
- Hardcopy of FFQ.

Data entering into online FFQ database

1. Click on '5yr FFQs'.
2. Click on 'New FFQ 5yr'.
3. Leave as blank for visit.
4. For study ID, key in ID as stated on FFQ hardcopy and 'ID**_home**' if FFQ is a home version, i.e. only frequencies and not amounts.
5. Key in details as found on hardcopy FFQ.
6. Click 'Create Ffq5yr'.
7. Follow the rules written in 'EAT5 FFQ codebook_6Oct17' excel.

If FFQ food item amount was not stated in the correct units

1. Use 'EAT5 FFQ codebook_6Oct17' excel.
2. Follow what is written under the different food item.
3. Record on hardcopy.

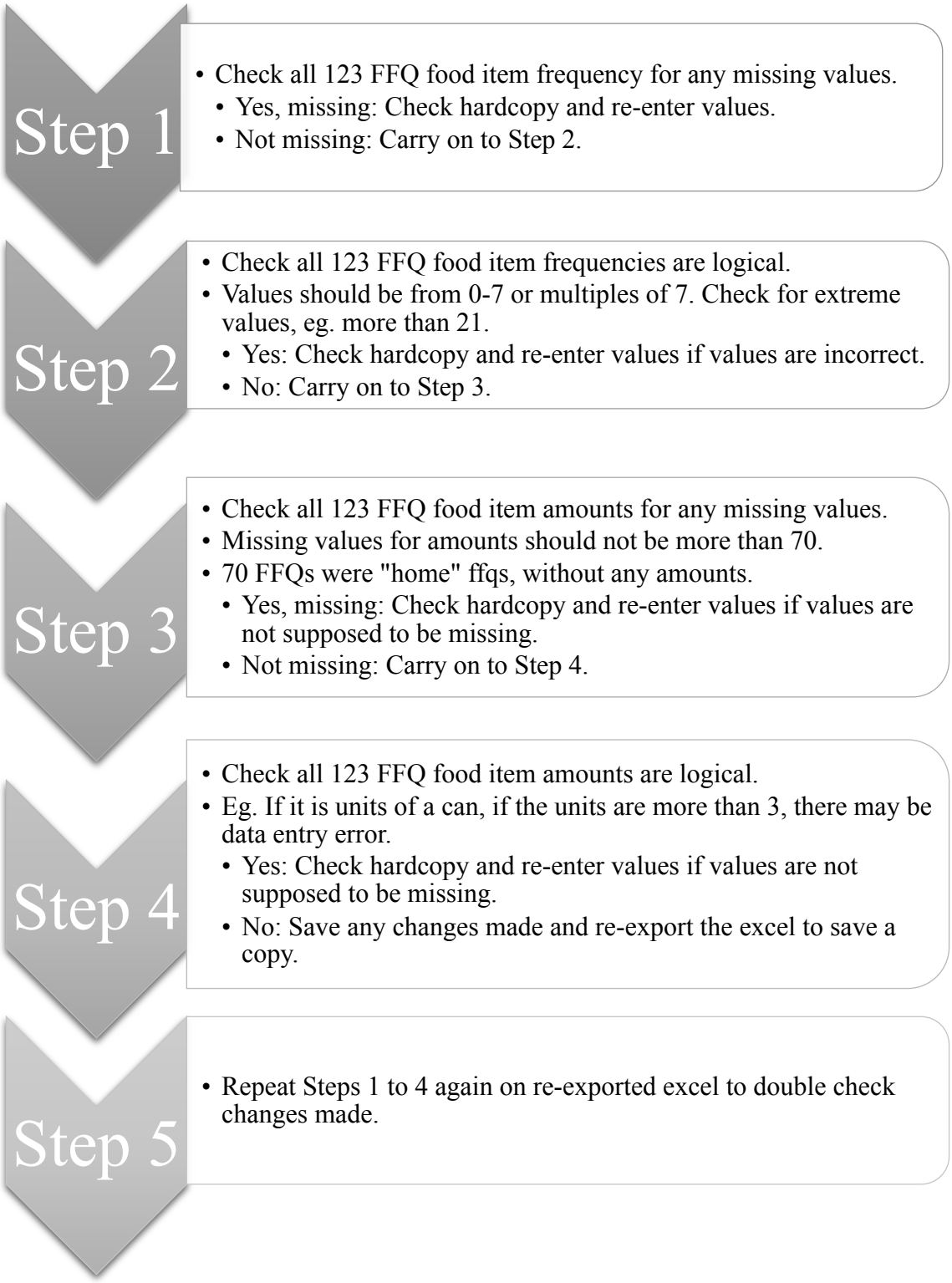
If there is missing value in FFQ hardcopy

1. Leave as blank if the value is missing in the FFQ hardcopy.

FFQ checking

1. Open 'https://ffq.otago.ac.nz/ffq5yrs'.
2. Click on 'Admin'.
3. Click on 'Export FFQ5'.
4. Save excel spreadsheet generated.
5. Import the excel spreadsheet into Stata 13 software.
6. Check the data in Stata 13 using the flowchart.

7. Edit the changes in the online FFQ database by choosing the FFQ and clicking on ‘edit’.



Appendix K

Flowchart for determining FFQ nutrient lines

The values of the nutrient lines were developed by Dr. Michelle Jospe with discussion with the Candidate and her supervisors ALH and RT. The Candidate developed the flowchart for determining the FFQ nutrient lines and checked through all the FFQ nutrient lines in the Excel output and after the nutrient lines had been put into the online FFQ database.

Example for: 'FFQ question 11. Cornflakes or rice bubbles'

Step 1: Look at frequency of consumption from previous studies.

Food description		2 year old Frequency	2 year old Percentage	10 year old	10 year old Percentage	Average (%)
cornflakes (skippy, sanitarium)		15	35%	5	56%	45%
rice bubbles, kelloggs		17	40%	1	11%	25%
ricies sanitarium		11	26%	3	33%	30%
	Total	43		9		

Step 2: Key into Kai-culator using the average percentages as grams.

Recipe: Q11 corn flakes or rice bubbles (1)

Moisture retention: 100 % Method Time: 0 mins Temp:

#	Food item	Amount	Id
1	Breakfast cereal, cornflakes toasted, ready to...	45.00 g	D1057
2	Rice bubbles kelloggs	25.00 g	D42
3	Ricies sanitarium	30.00 g	D1029

Step 3: In Kai-culator under 'records', create a new record and select the recipe and key in the amount eaten as 100g.

_	Diary Item	Time	Food Item	Amount
11	Q11		Q11 corn flakes or rice bubbles (1)	100.0 g

Step 4: Export the 'record' so that the nutrient line will be generated in excel with nutrients per 100g.

descriptor	WATER (g)	ENERC (kJ)	PROT (g)	FAT (g)	CHOAVL (g)	NA (mg)	FIBTG (g)
Q11 corn flakes or rice bubbles (1)	3.885	1488.945	6.828	1.190	78.168	799.000	2.275

Step 5: Calculate the density based on step 1 (frequency) and step 2 (density from each item from Kai-culator).

Using 'form density': Eg. from Rice bubbles kellogs D42 in Kai-culator.
The final density used for the FFQ food item would be 0.12.

Rice bubbles kelloggs

Measure	Value	Order
1 cup (250mL)	30	0
1 cup	30	1
Densities	-1	-1
Shape density (solid)	0.12	0
Form density	0.12	0

Frequency	Food Code	Density	Density * Frequency
45%	D1057	0.12	0.05
25%	D42	0.12	0.03
30%	D1029	0.13	0.04
		Sum:	0.12

Step 6: Calculate the nutrients from amount of food item eaten

Eg. In the EAT5 FFQ: Frequency is once a week; Amount eaten is 83.3mL each time.

11.	Cornflakes or rice bubbles	1	ml	83.3

Amount eaten in grams would be: 83.3 x 0.12 = 10.0g
Energy from cornflakes = (1488.945 kJ (from Step 4) / 100g) x 10.0g = **148.9 kJ.** Repeat this for all the other nutrients.

Appendix L

Decisions made to determine non-starch polysaccharide values

The decision of which non-starch polysaccharide (NSP) values to use for the FFQ nutrient lines and for the individual food items from the WDR was determined by the Candidate with advice from supervisors ALH and RT.

Example for EAT5 FFQ food items:

FFQ questions 51-54 do not have complete NSP values (columns D-F) from FOODfiles 2014. Values are obtained from FOODfiles 2010 from similar food items (columns N-P). Column Q gives the details of the food item used. FFQ question 55 has FOODfiles 2014 values for the NSP and hence the values are used.

A	B	C	D	E	F	N	O	P	Q
FFQ Question #	Variable label (n=123 items)	Fibre (g/100g)	Total PSACNS (g/100g)	PSACNS_soluble (g/100g)	PSACNS_Insoluble (g/100g)	Total PSACNS (g/100g)	PSACNS_soluble (g/100g)	PSACNS_Insoluble (g/100g)	Comments
	Reference	FOODfiles 2014	FOODfiles 2014	FOODfiles 2014	FOODfiles 2014	To USE	To USE	To USE	FOODfiles 2010: Item number, descriptor
51	Green cabbage	1.7947	0	0	0	1.6	0.7	0.9	X26, Cabbage,White,inner leaves,raw
52	Spinach or silverbeet	2.3264	0.0001	0.0001	0.0001	2.1	0.8	1.3	X116, Spinach,leaves,boiled,drained
53	Lettuce or salad leaves	1.56	1.54	0.0001	0.0001	3.04	1.89	1.15	X307, Lettuce,Cos
54	Cucumber	0.9	0.0002	0.0002	0.0002	0.6	0.2	0.4	X45, Cucumber,flesh,raw
55	Raw tomato	1.2	1.19	0.67	0.52	1.19	0.67	0.52	

Example for individual food items from WDRs:

The individual food items from the WDRs (1010 individual items in total) with FOODfiles 2014 values were exported from Kai-culator to excel. The following was checked: 1. Any blanks for total NSP (column E) if there is a fibre value (column D); 2. Insoluble (column G) and soluble (column F) NSP adds up to total NSP (column E). Any missing values from the above 2 will be checked and replaced by a value from similar foods from FOODfiles 2010 (changes highlighted in yellow and description of FOODfiles 2010 foodID used in column J.

A	B	C	D	E	F	G	J
FoodID	Food Name	EnergykJ/100g	FIBg/100g	PTg/100g	PSOLg/100g	PINg/100g	FOODfiles 2010
X1020	Broccoli, inflorescent vegetable, raw	112.53	3.4	3.8	1.6	2.2	
X1021	Broccoli, inflorescent vegetable, microwaved	121.48	3.4	3.2	1.3	1.9	
X1022	Broccoli, inflorescent vegetable, boiled, drained, no salt added	99.7	3.4	3.2	1.3	1.9	
X1030	Ginger, root vegetable, rhizomes, preserved	356.15	2.14	2	1	1	
X1045	Spinach, leafy vegetable, raw, English	57.3	2.2	2.1	0.8	1.3	X118
X1046	Spinach, leafy vegetable, boiled, drained, no salt added, English	62.61	2.3	2.1	0.8	1.3	X118
X1048	Capsicum, fruit vegetable, stalk & seeds removed, raw, Red	108.34	1.5	1.6	1.3	0.3	X272
X1049	Capsicum, fruit vegetable, stalk & seeds removed, raw, Green	72.64	1.5	1.6	1.3	0.3	X272

Appendix M

Flowchart for determining FFQ nutrient lines for fibre fractions

The values of the nutrient lines for fibre fractions were developed by the Candidate with discussion with supervisors ALH and RT. The Candidate developed the flowchart for determining the FFQ nutrient lines for fibre fractions and checked through all the FFQ nutrient lines in the Excel and after the nutrient lines had been put into the online FFQ database.

Example for: Resistant starch

Step 1: Look at whether there are similar food items from the resistant starch report, 2012

Example:

FFQ Question #	Description	Item name (from reference)	Resistant starch (g/100g)	Comments
16	White rice	Rice	0.37	Resistant starch report, 2012

Step 2: If there are no similar food items in the resistant starch report, look at the reference Murphy et al., 2008

Example:

FFQ Question #	Description	Item name (from reference)	Resistant starch (g/100g)	Comments
17	Brown rice	Rice, brown, cooked	1.70	Murphy et al., 2008

Step 3: If the FFQ food item is a mixture of multiple items, use percentage frequencies from nutrient line in **Appendix K**

Example:

Corn chips: 0.87g of resistant starch /100g food (17% frequency)

Potato chips: 0.21g of resistant starch /100g food (83% frequency)

FFQ Question #	Description	Item name (from reference)	Resistant starch (g/100g)	Comments
				Resistant starch report, 2012
88	Crisps, corn chips, corn snacks (eg. Cheezels)	Corn chips+Potato crisps	0.32	(0.87*0.17)+(0.21*0.83)

Example for: Hemicellulose, pectin, cellulose, klason lignin

Using EAT5 FFQ question 26. Pears

Step 1: Look at whether there are similar food items from the reference Marlett and Cheung, 1997

FFQ Question #	Variable label	Item name (from reference)
26	Pears	Pear, unpeeled

Step 2: Convert the values to per 100g basis. The values from Marlett and Cheung, 1997 are per serving basis.

Per serve of pear:

Weight/serve	Total fiber (g/serve)	Hemicellulose soluble (g/serve)	Pectin soluble (g/serve)	Cellulose insoluble (g/serve)	Hemicellulose insoluble (g/serve)	Pectin insoluble (g/serve)	Klason lignin insoluble (g/serve)
166	4.6	0.3	0.4	1.2	1.4	0.7	0.6

to per 100g by dividing by 166g and multiply by 100g to get:

Total fiber (g/100g)	Hemicellulose soluble (g/100g)	Pectin soluble (g/100g)	Cellulose insoluble (g/100g)	Hemicellulose insoluble (g/100g)	Pectin insoluble (g/100g)	Klason lignin insoluble (g/100g)
2.8	0.2	0.2	0.7	0.8	0.4	0.4

Step 3: Check that the fibre value (g/100g) is close to the fibre value from FOODfiles 2014.

Total fibre (g/100g) Murphy	Total fibre (g/100g) FOODfiles	Difference
2.8	3.0	-0.2

Appendix N

Statistical analysis plan & lessons learnt for study 3

The Candidate developed the statistical analysis plan with advice from supervisors JH, ALH and RT. References are not included as they can be found in the main body of the thesis.

Statistical analysis plan for POI5micro paper

Version: 15 Dec 2018

Title: Gut microbiota and diet relationships in young children and their association with overweight and obesity.

Primary aim: To determine associations between diet and the gut microbiota in children 5 years of age.

Secondary aim: To determine the extent to which associations between diet and obesity explain associations between the gut microbiota and obesity at 5 years of age.

What we know:

- Diet is one of the main contributors to modulation of the gut microbiota (studies in adults).
    - Dietary patterns: plant vs animal based diets, dietary fats and proteins, carbohydrates and fibre, non-starch polysaccharides, cellulose and hemi-cellulose, whole grains (adults: review) (Sheflin, Melby, Carbonero, & Weir, 2017).
    - Westernized diet type had a greater impact (based on effect size) compared to body mass index on causing dysbiosis of the gut microbiota (Davis, Yadav, Barrow, & Robertson, 2017).
    - Staphylococcus spp. positively associated with energy intake (Bervoets et al., 2013).
    - *Bifidobacteria* growth is stimulated by dietary oligosaccharides derived from plant and milk (Meyer & Stasse-Wolthuis, 2009)
- Diet is one of the main contributors to modulation of the gut microbiota (studies in infants).
    - (7-12mos) Dietary fibre and intake of fruit and vegetables and alpha diversity (Leong et al., 2018).
- Gut microbiota is different between obese and normal BMI children.
    - Firmicutes and Bacteroidetes ratio (F/B) higher in the obese cohort, ages 3-18 year old (Bervoets et al., 2013; Hou et al., 2017).
    - However, high Lactobacillus spp. was associated with obese microbiota (Bervoets et al., 2013).
    - Increased representation of members of the Firmicutes in individuals who became overweight by 7 years of age (Kalliomaki, Collado, Salminen, & Isolauri, 2008).
    - Fimicutes (Clostridia > Ruminococcus) associated with genes involved in carbohydrate catabolism (Turnbaugh & Gordon, 2009).
    - Bacteroidetes (Prevotella and Bacteroides spp.) efficient utilizers of plant polysaccharides, linked with diminished body mass, high in children from rural African villages (De Filippo et al., 2010; Kovatcheva-Datchary, Tremaroli, & Bäckhed, 2013).
- Gut microbiota is different between obese and normal BMI adults.
    - Faecalibaterium prausnitzii negatively associated with inflammatory markers that can alleviate obesity (adults: review) (Tremaroli & Bäckhed, 2012).
- Gut microbiota is different between obese and normal BMI (animal model).

    - Lactobacillus (different species) associated with lower/ reduced body weight (mice: review) (Harakeh et al., 2016).
- Association of diet + obesity + gut microbiota in children
    - (13-16 yr) Dietary pattern with plant and fermented foods associated with higher proportions of Bacteroides (Bacteroidaceae) and Bifidobacterium (Bifidobacteriaceae-Actinobacteria), lower Prevotella (Prevotellaceae) compared to western dietary pattern. Increase of prevotella (Prevotellaceae) and decrease of Bacteroides (Bacteroidaceae) and Ruminococcaceae had a higher risk of obesity (Jang, Choi, Kang, Park, & Lee, 2017).

What we want to know (Research questions):

- The food we eat affects both the gut microbiota and our health. Many studies have shown that those who are obese have a dysbiosis of the gut microbiota and differences in the gut microbiota bacterial taxa from those who are lean. Could one of the main causes of this dysbiosis be due to diet that subsequently leads to obesity? It is especially important to look at this problem in young children because they are still growing rapidly.

How to find out the answers:

Table 1: Demographic characteristics of study participants (with and without full dataset).

- Sex
- Ethnicity
- Age
- Mode of delivery
- Maternal parity
- Maternal education
- Maternal BMI
- Household deprivation
- Randomised group

Figure 1: Gut microbiota profiles using principal component analysis of genera as compositional data.

Table 2: Correlations of baseline, early diet, nutrient, food groups and body composition with alpha diversity and gut microbiota profiles.

Baseline

- Child birth weight, g
- Parity
- Maternal education
- Maternal pre-pregnancy BMI, kg/m2
- Household deprivation
- Mode of delivery

Early diet

- Exclusively breastfed, wk
- Any breastfeeding, wk
- Age at introduction of solids, wk

Nutrient

- Energy, kJ

- Carbohydrate, g
- Fat, g
- Protein, g
- Total fibre, g
- Total NSP, g
- Insoluble NSP, g
- Soluble NSP, g

Food group

- Higher fibre more healthy cereals, g
- Lower fibre more healthy cereals, g
- Higher fibre less healthy cereals, g
- Lower fibre less healthy cereals, g
- Nuts, seeds, legumes, g
- Fruits, g
- Vegetables, g
- Potatoes, g
- Dairy, g
- Yoghurt, g
- Meat, fish, poultry, g

Body composition at 5 y

- BMI z-score
- DXA fat3, %
- FMI z-score3

Table 3: Regression models of baseline, early diet, nutrient, food groups and body composition with alpha diversity and gut microbiota profiles. (With adjustment for covariates in the models)

Lessons learnt for the Candidate:

- It is of particular importance to form initial research questions. A challenge with big data is the temptation to do numerous exploratory analyses until something interesting in the form of statistical significance appears. This is analogous to a fishing expedition and is not good research practice.
- Importance of explaining figures and statistics used. It is important when developing figures that the researcher explains the figures in a way that would be understandable to researchers from different fields. This is because most scientific areas now emphasize multi-disciplinary approaches and hence for research to be applied in different areas, figures shown should be explained, and statistics used to develop the figures should be reported so that other researchers may benefit from using similar statistics if their data allows for that.
- In addition to similar points from study 1, it is also important to know what analyses is being carried out in the background. Bioinformatics software programmes such as STAMP are useful as they are easy to use and able to produce good quality figures. However, the statistical tests are limited to what the program provides; the user is unable to transform the data in the program; and unable to adjust for confounders; and the user has to create groups in the metadata before loading into the program if need to compare between groups. The figure below shows the exploration of the POI5 microbiota and dietary data using STAMP.

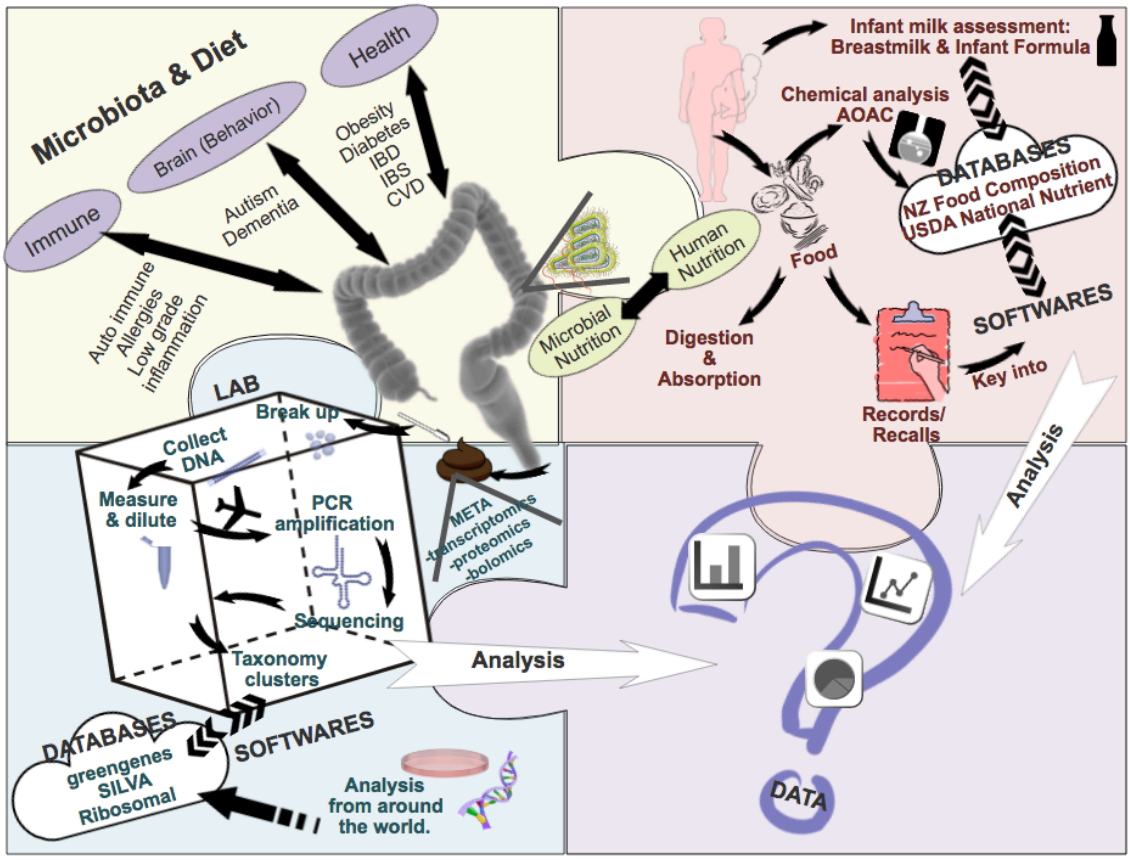
Appendix O

Final thoughts

The microbiota and diet area to the Candidate is analogous to a puzzle (see figure over page), where the pieces of the puzzle need to be collated to be able to solve health problems. The thesis contributes to the literature by:

- Identifying a gut microbiota profile associated with BMI z-score in 5 year NZ children (yellow puzzle piece)
- Identifying 3 gut microbiota profiles in 5 year NZ children, and showing that gut microbiota differs between different methods of complementary feeding (blue puzzle piece).
- Validating an FFQ for nutrients of relevance to the gut microbiota (pink puzzle piece).
- Demonstrating the usefulness of mediation analyses, compositional principal component analysis and regression models in this area (purple puzzle piece).

Candidate used the SmartDraw software for graphical design of this figure. Images used in the figure were sourced from free icons and images websites (www.flaticon.com and www.smart.servier.com).



Pictorial summary of key interrelationships between the diet and the gut microbiota that will be discussed in this thesis. **Yellow:** Shows the importance of studying the gut microbiota. **Blue:** Shows the steps for microbiota analysis. **Pink:** Shows the steps for dietary analysis. **Purple:** Shows how diet and microbiota analysis need to be analysed together.